

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N	A2	(11) International Publication Number: WO 99/64562 (43) International Publication Date: 16 December 1999 (16.12.99)
(21) International Application Number: PCT/US99/13266 (22) International Filing Date: 11 June 1999 (11.06.99) (30) Priority Data: 60/089,049 12 June 1998 (12.06.98) US 60/089,050 12 June 1998 (12.06.98) US (71) Applicant: THE PENNSYLVANIA STATE UNIVERSITY [US/US]; 113 Technology Center, 200 Innovation Boulevard, University Park, PA 16802-7000 (US). (72) Inventors: GUILTINAN, Mark, J.; 801 Crabapple Court, State College, PA 16803 (US). KIM, Kyung-Nam; 6023 Avila Street, El Cerrito, CA 94530 (US). (74) Agent: MONAHAN, Thomas, J.; The Pennsylvania State University, 113 Technology Center, 200 Innovation Boulevard, University Park, PA 16802-7000 (US).		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: EXPRESSION CONTROL ELEMENTS FROM GENES ENCODING STARCH BRANCHING ENZYMES		
(57) Abstract Nucleic acid molecules for directing and modulating gene expression in transformed cells are disclosed. The molecules comprise 5' and 3' regulatory sequences isolated from genomic clones having homology in their coding regions with maize genes encoding Type I and Type IIB starch branching enzymes. Chimeric genes comprising these expression controlling nucleic acid molecules, and cells transformed with such chimeric genes, are also disclosed.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

**EXPRESSION CONTROL ELEMENTS FROM
GENES ENCODING STARCH BRANCHING ENZYMES**

Pursuant to 35 U.S.C. §202(c), it is
acknowledged that the U.S. Government has certain rights
in the invention described herein, which was made in part
with funds from the Department of Energy, Grant No. DE-
5 FG02-96ER20234.

This application claims priority to U.S.
Provisional Application Serial Nos. 60/089,049 and
60/089,050, filed June 12, 1998, the entireties of which
are incorporated by reference herein.

10

FIELD OF THE INVENTION

The present invention relates to the field of
genetic manipulation in plants. In particular, the
invention provides a novel transcription and translation
15 control elements isolated from genes encoding starch
branching enzymes.

BACKGROUND OF THE INVENTION

Various publications or patents are referenced
20 in this application to describe the state of the art to
which the invention pertains. Each of these publications
or patents is incorporated by reference herein.

Starch provides carbon and energy for
vegetative and reproductive development of most higher
25 plants. It is found as a water-insoluble granule which
is mainly composed of two different polysaccharides,
amylose and amylopectin. Amylose is considered to
consist of linear α -1,4-linked glucose chains of about
1,000 residues long. However, amylopectin is a more
30 highly branched macromolecule consisting of linear α -1,4-

- 2 -

glucose chains with α -1,6-glucosidic bonds at branch points.

Starch is synthesized in higher plants through the action of four classes of enzymes, ADP-glucose pyrophosphorylase (EC 2.7.7.23), starch synthase (EC 2.4.1.21), starch branching enzyme (EC 2.4.1.28) and starch debranching enzyme (EC 3.2.1.41). Starch branching enzymes (SBE) catalyze the formation of α -1,6 glucan linkages, thereby playing an important role in the synthesis of the amylopectin fraction of starch.

Multiple forms of starch branching enzymes (SBE) differing in enzymatic and biochemical properties have been identified and characterized in various plants. Three SBE isoforms, SBEI, IIa and IIb, have been isolated and characterized in maize. Isolation of the maize cDNAs encoding the SBE isoforms (*Sbe1*, *2a* and *2b*) enabled the investigation of the *Sbe* genes at molecular level (Fisher et al., Plant Physiol. 102: 1045-1046, 1993; Fisher et al., Plant Physiol. 108: 1313-1314, 1995; Gao et al., Plant Mol. Biol. 30: 1223-1232, 1996; Plant Physiol. 114: 69-78, 1997). Fisher et al. (Plant Physiol. 110: 611-619, 1996) determined that SBEIIa and IIb are the product of separate genes, and Gao et al. (1996, 1997, *supra*) demonstrated that the *Sbe* genes are differentially expressed during kernel development and in various tissues, suggesting that they play distinct roles in starch biosynthesis. For example, while *Sbe1* and *2a* are expressed in vegetative tissues, *Ae* is not. Moreover, unlike *Sbe1* and *2b*, *Sbe2a* is more highly expressed in embryos than in endosperm. However, all three genes are highly expressed in developing maize seed from mid to late development in both embryo and endosperm tissue.

Developmental and tissue specificity of gene

- 3 -

expression is governed by the expression regulatory regions of a gene. Typically, these include promoters, enhancers, 5' untranslated leaders, translation terminators and polyadenylation signals, among others.

- 5 These *cis*-acting sequences control gene expression by regulating the timing, location and amount of transcription of the gene and/or stability or efficiency of translation of the encoded mRNA.

- As maize seed is a major global food source,
10 the ability to engineer protein production in seed will have many applications in agriculture and industry. High-level and tissue-specific expression regulatory sequences will be needed for the development of such genetically engineered plants. The expression regulatory
15 sequences of the above-described *Sbe* and *Ae* genes from maize would be well suited for such applications, but heretofore have been unavailable.

SUMMARY OF THE INVENTION

- 20 According to one aspect of the present invention, an isolated nucleic acid molecule for controlling expression of genes in transformed plant cells is provided, which comprises a segment of an *Sbe1* gene from *Zea mays* or related species. In a preferred
25 embodiment, the nucleic acid molecule is isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:1. In one embodiment, the segment begins at a location about 3,000 bases upstream from the
30 transcription initiation site of the gene, and ends at a location about 250 bases downstream from the transcription initiation site. The molecule contains one or more specific regions for effecting high level

- 4 -

expression or sugar regulation. In another embodiment, the segment comprises a 3' untranslated region commencing at a stop codon for the gene's coding sequence, and ending at a location about 5900 bases downstream from the gene's transcription initiation site.

According to another aspect of the invention, a DNA segment for effecting expression of coding sequences operably linked to the segment is provided, which is isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1. Preferably, the gene is a maize *Sbe1* gene. In one embodiment, the segment comprises a promoter and a transcription initiation site, and may further comprise (1) an element included in the first exon of SEQ ID NO:1, which is capable of increasing promoter activity of homologous or heterologous promoters operably linked thereto, or (2) an element that confers sugar-regulatability on expression of the coding sequence. In another embodiment, the segment comprises a 3' untranslated region of the gene.

According to another aspect of the invention, another isolated nucleic acid molecule for controlling expression of genes in transformed plant cells is provided, which comprises a segment of a plant *Ae* gene. In a preferred embodiment, the nucleic acid molecule is isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:2, and most preferably it is a maize gene. In one embodiment, the segment begins at a location about 3,000 bases upstream from a transcription initiation site of the gene, and ends at a location about 100 bases downstream from the transcription initiation site. In another embodiment, the segment comprises a 3'

- 5 -

untranslated region commencing at a stop codon for gene's coding sequence, and ending at a location about 20,500 bases downstream from the gene's transcription initiation site.

5 In another aspect of the invention, a DNA segment for effecting expression of coding sequences operably linked to the segment is provided, which is isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by
10 exons of SEQ ID NO:2. Preferably, the gene is a maize *Ae* gene. In one embodiment, the segment comprises a promoter and a transcription initiation site. In another embodiment, the segment comprises a 3' untranslated region.

15 According to another aspect of the invention, a chimeric gene comprising a coding sequence operably linked to one or more of the aforementioned expression-regulatory sequences is provided. The chimeric gene preferably is inserted into a vector for transforming
20 cells. Cells transformed with the vector are provided. In a preferred embodiment, they are plant cells, and are regenerated into fertile transgenic plants.

 These and other features and advantages of the present invention will be described in greater detail in
25 the description and examples set forth below.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Structure of the lambda clone 5-1-1 containing the *Sbe1* gene. **Fig. 1A.** Restriction map of
30 the 5-1-1 clone. pBI5-1 and pBI5-2 indicate subclones in plasmid pBluescript SK⁻. **Fig. 1B.** Genomic structure of the *Sbe1* gene. The thin black lines indicate the 5'- or 3'-flanking sequences of the *Sbe1* gene. The solid black

- 6 -

boxes indicate exons and the open boxes denote introns. The numbers represent positions relative to the transcription initiation site (+1).

Figure 2. Nucleotide sequence of the *Sbe1* gene and 5'- and 3'-flanking regions. Flanking regions and introns are shown in lowercase letters, while exons are presented in uppercase letters. The deduced amino acid sequences are shown below the string of exon sequences. Numbers indicate the distance relative to the transcription start site (+1) which is indicated by the arrowhead. The consensus TATA and G-box sequences as well as putative polyadenylation signal are underlined. The regions containing at least 82% sequence homology with the rice *Sbe1* 5' flanking region is also underlined. The asterisk and dot indicate the stop codon and putative polyadeylation site, respectively.

Figure 3. Genomic structure of the *Ae* gene. The complete structure of the *Ae* gene was constructed using two overlapping genomic clones: λ 3-2-1 and 7-2-1. The thick black lines indicate the 5'- or 3'-flanking sequences of the *Sbe1* gene. The solid -black boxes indicate exons and the open boxes denote introns. The position of a putative TATA-box relative to the transcription initiation site (+1) is indicated.

Figure 4. Nucleotide sequence of the *Ae* gene and 5'-and 3'-flanking regions. Numbers indicate the distance relative to the transcription start site (+1) which is indicated by the arrow. Intron sequences are omitted. Putative *cis*-elements are indicated below the sequence and boxed. The end points of the 5' deletion constructs were indicated by arrowheads below the first nucleotides of the deletions. Direct repeat sequences are underlined. The translation initiation codon (ATG)

- 7 -

and the translation termination codon (TGA) are double underlined. The dot indicates polyadenylation site. The region exhibiting 80% sequence similarity with the PREM-2 internal and 3' LTR region is boxed. The wavy line
5 represents polypurine tract.

Figure 5. Schematic diagram of the 5' deletion chimeric constructs. **Fig. 5A.** The thick black lines denote the Ae promoter sequences. The numbers at left indicate deletion-end points relative to the
10 transcription initiation site (+1) of the Ae gene. The open and stripped boxes indicate luciferase gene and nopaline synthase 3' end sequences, respectively. **Fig. 5B.** The junction sequences between the Ae gene and LUC. The BamHI site used to join the two genes is underlined.
15 The translation start site of LUC is indicated by boldface letters. **Fig. 5C.** Effect of 5' deletions on Ae promoter activity. The relative activity values of the constructs are percentages of pKL201 level. Each value represents the average of three independent shootings.
20 Error bars indicate standard errors of the means.

Figure 6. Effect of *SbeI* gene exon/introns and 3' end on the level of LUC expression driven by the *SbeI* promoter. **Fig. 6A.** Schematic diagram of chimeric *SbeI* promoter-luciferase constructs. Numbers indicate
25 distance relative to the *SbeI* transcription start site. Translation initiation starts at a position, + 28. The light grey (stippled) boxes indicate the *SbeI* promoter region. Angled lines indicate exons and introns in the *SbeI* gene. Open and solid black boxes indicate
30 luciferase (LUC) reporter gene and nopaline synthase 3' end sequences, respectively. The striped box indicates the *SbeI* 3' flanking sequence. **Fig. 6B.** The junction sequences between the *SbeI* gene and LUC. The BamHI sites

- 8 -

used to join the two genes are underlined. The translation start site of *LUC* is indicated by boldface letters. **Fig. 6C.** Expression levels of the construct shown in (A). *LUC/GUS* ratios were calculated as
5 described in Methods. Each value represents the average of four independent shootings. Error bars indicate standard errors of the mean.

Figure 7. Effect of 5' deletions on *SbeI* promoter activity. **Fig. 7A.** Schematic diagram of the 5'
10 deletion chimeric constructs. The thick black lines denote the *SbeI* promoter sequences. The numbers at left indicate deletion-end points relative to the transcription initiation site (+1) of the *SbeI* gene. The light grey (stippled) boxes and the thin black angled
15 line represent the first exon and intron in the *SbeI* gene, respectively. The open boxes indicate the luciferase gene. The solid black boxes denote nopaline synthase 3' end sequences. **Fig. 7B.** The relative activity levels of the constructs shown in (A). The
20 relative activity values are percentages of pKLN101 level. Each value represents the average of six to eight independent experiments. Error bars indicate standard errors of the means.

Figure 8. Linker-scan analyses of the 60-bp
25 region in the *SbeI* promoter. **Fig. 8A.** Schematic diagram of the linker-scan constructs. DNA sequence of the 60-bp region in the *SbeI* promoter is shown to the right of the wild-type construct pKLN105. The mutated bases in the linker-scan constructs are shown in lowercase letters.
30 Dashes represent the unaltered nucleotides. For an explanation of the other symbols, refer to the legend to Figure 13. **Fig. 8B.** Relative *LUC* activity levels of the constructs shown in (A). The relative activity values

- 9 -

are percentages of construct -314 level. Each value represents the average of four independent experiments. Error bars indicate standard errors of the means.

Figure 9. Sucrose responsiveness of the *Sbe1* promoter. Each construct was bombarded onto the maize endosperm suspension cell cultures supplemented with 0% (-) sucrose or 9% (+) sucrose and incubated for 48 hr at 25°C in the dark. The relative activity values are percentages of pKLN101 level in 0% sucrose. Each value represents the average of three independent experiments. Error bars indicate standard errors of the means.

Figure 10. Effect of mEmBP-1 overexpression on *Sbe1* promoter activity. 4 µg each of reporter plasmid (*Sbe1* promoter-LUC; pKLN101, or Ubiquitin-LUC; pACH18) and reference plasmid (CaMV 35S-GUS; pBI221) were co-precipitated onto gold particles with or without 4 µg of an effector plasmid (CaMV 35S-mEmBP-1). Maize endosperm suspension cells were bombarded with the gold particles and incubated at 25°C for 24 hr in the dark. The relative activity values are percentages of the pKLN101 or pACH18 levels without mEmBP-1 overexpression. Each value represents the average of two independent shootings. Error bars indicate standard errors of the means.

25

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Various terms used throughout the specification and claims to describe the invention. Unless otherwise specified, these terms are defined as set forth below.

With reference to nucleic acid molecules, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that

- 10 -

is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may
5 comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule.

With respect to RNA molecules, the term
10 "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its
15 natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the
20 compound of interest (e.g., nucleic acid, oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the
25 compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

Nucleic acid sequences and amino acid sequences can be compared using computer programs that align the
30 similar sequences of the nucleic or amino acids thus define the differences. For purposes of this invention, the GCG Wisconsin Package version 9.1, available from the Genetics Computer Group in Madison, Wisconsin, and the

- 11 -

default parameters used (gap creation penalty=12, gap extension penalty=4) by that program are the parameters intended to be used herein to compare sequence identity and similarity.

5 The term "substantially the same" refers to nucleic acid or amino acid sequences having sequence variation that do not materially affect the functionality of *cis* acting regulatory sequences (e.g, promoters, transcriptional response elements, etc.) or the nature of
10 the encoded gene product (i.e. the structure, stability characteristics, substrate specificity and/or biological activity of the protein). With particular reference to nucleic acid sequences, the term "substantially the same" is intended to refer to conserved sequences governing
15 expression and to the coding region (referring primarily to degenerate codons encoding the same amino acid, or alternate codons encoding conservative substitute amino acids in the encoded polypeptide).

 The terms "percent identical" and "percent
20 similar" are also used herein in comparisons among nucleic acid sequences. When referring to nucleic acid molecules, "percent identical" refers to the percent of the nucleotides of the subject nucleic acid sequence that have been matched to identical nucleotides by a sequence
25 analysis program. When referring to amino acid sequences, "percent identical" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical amino acids in the compared amino acid sequence by a sequence analysis program. "Percent
30 similar" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical or conserved amino acids. Conserved amino acids are those which differ in structure but are similar

- 12 -

in physical properties such that the exchange of one for another would not appreciably change the tertiary structure of the resulting protein. Conservative substitutions are defined in Taylor (1986, J. Theor. Biol. 119:205).

With respect to oligonucleotides or other single-stranded nucleic acid molecules, the term "specifically hybridizing" refers to the association between two single-stranded nucleic acid molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

A "coding sequence" or "coding region" refers to a nucleic acid molecule having sequence information necessary to produce a gene product, when the sequence is expressed.

The term "operably linked" or "operably inserted" means that the regulatory sequences necessary for expression of the coding sequence are placed in a nucleic acid molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement other transcription control elements (e.g. enhancers) in an expression vector.

When describing the organization of a nucleic acid molecule, the term "upstream" refers to the 5' direction and the term "downstream" refers to the 3'

- 13 -

direction.

Transcriptional and translational control sequences, sometimes referred to herein as "expression control" sequences or elements, or "expression
5 regulating" sequences or elements, are DNA regulatory elements such as promoters, enhancers, ribosome binding sites, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell. The term "expression" is
10 intended to include transcription of DNA and translation of the mRNA transcript.

The terms "promoter", "promoter region" or "promoter sequence" refer generally to transcriptional regulatory regions of a gene, which may be found at the
15 5' or 3' side of the coding region, or within the coding region, or within introns. Typically, a promoter is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The typical 5' promoter
20 sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter
25 sequence is a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A "vector" is a replicon, such as plasmid,
30 phage, cosmid, or virus to which another nucleic acid segment may be operably inserted so as to bring about the replication or expression of the segment.

The term "nucleic acid construct" or "DNA

- 14 -

construct" is sometimes used to refer to a coding sequence or sequences operably linked to appropriate regulatory sequences and inserted into a vector for transforming a cell. This term may be used

5 interchangeably with the term "transforming DNA". Such a nucleic acid construct may contain a coding sequence for a gene product of interest, along with a selectable marker gene and/or a reporter gene.

A "heterologous" region of a nucleic acid
10 construct is an identifiable segment (or segments) of the nucleic acid molecule within a larger molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA
15 that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences
20 having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

A cell has been "transformed" or "transfected"
25 by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. For example, the transforming DNA may be maintained on an episomal element such as a plasmid.
30 With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the

- 15 -

eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis.
5 A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

II. Description

The genes encoding starch branching enzymes
10 (SBEI, IIa and IIb) in maize are differentially regulated in tissue specificity and during kernel development. The expression-controlling sequences governing the differential regulation of genes encoding SBEI and SBEIIb are the subject matter of the present invention. These
15 sequences are of great utility and value in genetic modification of plants to produce other gene products in the same tissue- and developmentally specific manner as are the two aforementioned starch branching enzymes. This is accomplished by placing the coding sequence of a
20 gene of interest under the control of one or more of these expression-controlling sequences.

cdNAs for the starch branching enzymes have been disclosed, but such sequences, for the most part, do not contain the regulatory elements that control
25 expression of a gene. To obtain these expression-controlling sequences, the actual genes must be isolated. In accordance with the present invention, the isolation and characterization of two starch branching genes from maize, *Sbe1* and *Ae*, has enabled the identification of
30 elements that regulate the expression of these genes.

The inventors have isolated and sequenced a maize genomic DNA (-2190 to +5929) which contains the entire coding region of SBEI (*Sbe1*) as well as 5'-and 3'-

- 16 -

flanking sequences. Sequence analysis of this gene (SEQ ID NO:1) is shown in Figure 2. Using this clone, the complete genomic organization of the maize *Sbe1* gene was established. The transcribed region consists of 14 exons and 13 introns, distributed over 5.7 kb. A consensus TATA-box and a G-box containing a perfect palindromic sequence, CCACGTGG, were found in the 5'-flanking region. Genomic Southern blot analysis indicated that two *Sbe1* genes with divergent 5'-flanking sequences exist in the maize genome, indicating that they may be differentially regulated. A chimeric construct containing the 5'-flanking region of *Sbe1* (-2190 to +27) fused to the β -glucuronidase gene (pKG101) showed promoter activity after it was introduced into maize endosperm suspension cells by particle bombardment. Although the 2.2-kb 5'-flanking sequence between -2190 and +27 relative to the transcription initiation site was sufficient to promote transcription, addition of the transcribed region between +28 and +228 containing the first exon and intron resulted in high-level expression in maize endosperm suspension cells. A series of 5' deletion and linker-substitution mutants identified two critical positive cis-elements, -314 to -295 and -284 to -255. Electrophoretic mobility shift assay showed that nuclear proteins prepared from maize kernels interact with the 60-bp fragment containing these two elements. Expression of the *Sbe1* gene is regulated by sugar concentration in cultured maize endosperm suspension cells, and the region -314 to -145 is essential for this effect. Expression of mEmBP-1, a bZIP transcription activator, in maize endosperm suspension cells resulted in a 5-fold decrease in the *Sbe1* promoter activity, suggesting a possible regulatory role of the G-box present in the *Sbe1* promoter

- 17 -

from -227 to -220.

A genomic clone of a *Sbel* gene from rice has been described (Kawasaki et al., 1993, *supra*), as has an *Sbel* promoter from wheat (U.S. Patent No. 5,866,793 to Baga, et al.) However, the presently described expression regulatory elements have several distinguishing features, described in greater detail below, which have not been reported for the rice or wheat promoter elements.

10 The *amylose-extender* (Ae) gene encoding SBEIIb in maize is predominantly expressed in endosperm and embryos during kernel development. To obtain the expression-controlling sequences of this gene, a maize genomic DNA fragment (-2,964 to +20,485) containing the
15 Ae gene was isolated and sequenced. Sequence analysis of this gene (SEQ ID NO: 2) is shown in Figure 9. The maize Ae mRNA is derived from 22 exons distributed over 16,914 bp. Twenty one introns, differing in length from 76 bp to 4,020 bp, all have conserved junction sequences (GT .
20 . AG). Sequence analysis of the 5'- and 3'-flanking regions revealed a consensus TATA-box sequence located 28 bp upstream of the transcription initiation site as determined by primer extension analysis, and a putative polyadenylation signal observed 29 bp upstream of the
25 polyadenylation site based on cDNA sequence. Genomic Southern blot analysis revealed that a single Ae gene is present in the maize genome. Promoter activity was confirmed by testing a transcriptional fusion of the Ae 5'-flanking region between -2,964 and +100 to a
30 luciferase reporter gene in a transient expression assay using maize endosperm suspension cultured cells. 5' deletion analysis revealed that the 111-bp region from -160 to -50 is essential for high-level promoter activity.

- 18 -

The present invention is drawn to all expression controlling elements of genes encoding SBEI and SBEIIB, particularly as exemplified by the maize *Sbe1* and *Ae* genes described in detail herein. With respect to
5 the the maize *Sbe1* gene, the following sequences are considered particularly preferred for use in the present invention:

1. The 5' untranslated region, nucleotides - 2190 to +27 relative to the transcription initiation
10 site, which contains sequences sufficient to promote transcription;

2. The region containing the first exon and intron, nucleotides +28 to +228 relative to the transcription initiation site, which, when added to the
15 promoter region, results in high level gene expression in the endosperm;

3. The region between -314 and -145 relative to the transcription initiation site, which is needed for sugar regulation of gene expression; and within this
20 region

4. two critical positive cis-elements, -314 to -295 and -284 to -255; nuclear proteins prepared from maize kernels interact with the 60-bp fragment containing these two elements;

25 5. The 3' untranslated region, nucleotides +5390 to +5910, containing the polyadenylation signal and other sequences that may contribute to the stability or translation efficiency of the encoded mRNA.

With respect to the the maize *Ae* gene, the
30 following sequences are considered particularly preferred for use in the present invention:

1. the 5' untranslated region, nucleotides - 2,964 to +100 relative to the transcription initiation

- 19 -

site, which contains sequences sufficient to promote transcription, and within this region;

2. the 111-bp region from -160 to -50, which is essential for high-level promoter activity; and

5 3. The 3' untranslated region, nucleotides +16,695 to +20,485, containing the polyadenylation signal and other sequences that may contribute to the stability or translation efficiency of the encoded mRNA.

Although the genomic clones of the maize *Sbe1* and *Ae* are exemplified herein, this invention is intended to encompass nucleic acid sequences from other organisms, particularly higher plants, that are sufficiently similar to be used instead of the maize *Sbe1* and *Ae* nucleic acid molecules for the purposes described below. These include, but are not limited to, allelic variants and natural mutants of SEQ ID NOS: 1 and 2 and the particular portions thereof listed above, which are likely to be found in different maize cultivars, as well as in different plant species. Of particular relevance to this invention are monocotyledenous plant species, and most especially those of agronomic importance, such as wheat, rice, barley and sorghum, for example. Also of particular relevance to this invention are taxonomic relatives of *Zea mays* (which includes cultivated maize and teosinte), including other members of the genus *Zea*, such as *Zea diploperennis* (diploperennial teosinte), *Zea luxurians* and *Zea perennis* (perennial teosinte), as well as *Tripsacum spp.* and *Sorghum spp.*

Variants from other cultivars and species are expected to possess certain differences in nucleotide sequence. Moreover, it is known that 5' and 3' regulatory regions of genes, while often sharing overall functional similarities, do not share a high degree of

- 20 -

sequence homology. Accordingly, this invention provides nucleic acid molecules comprising one or more the above-described expression regulatory sequences, which are isolated from genes with coding sequences (exons) having
5 at least about 60% (preferably 70%, more preferably over 80% and most preferably over 90%) sequence homology with the respective coding sequences (exons) of the maize *Sbe1* or *Ae* genes, as exemplified by SEQ ID NO:1 or SEQ ID NO:2. The isolation of genes with exons having these
10 levels of homology is accomplished by nucleic acid hybridization at a selected stringency, which is described in greater detail below.

It will also be understood by persons skilled in the art that the *Sbe1* and *Ae* expression-regulating
15 elements of the invention include single- or double-stranded DNA or RNA.

The following description sets forth the general procedures involved in practicing the present invention. To the extent that specific materials are
20 mentioned, it is merely for purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning procedures, such as those set forth in Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989) (hereinafter "Sambrook et al.")
25 or Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1999) (hereinafter "Ausubel et al.") are used.

Nucleic acid molecules comprising the promoters and/or other expression regulatory sequences of the
30 invention may be prepared by two general methods: (1) they may be synthesized from appropriate nucleotide triphosphates, or (2) they may be isolated from biological sources. Both methods utilize protocols well

- 21 -

known in the art.

The availability of nucleotide sequence information, such as the aforementioned expression-controlling regions of SEQ ID NO:1 and SEQ ID NO:2, enables preparation of isolated nucleic acid molecules of the invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramidite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. Variants of the aforementioned sequences also may be synthesized as described above.

Sbe1 and *Ae* genes and their expression-regulatory elements also may be isolated from appropriate biological sources using methods known in the art. In the exemplary embodiment of the invention, the genomic clones having SEQ ID NO:1 and SEQ ID NO:2 were isolated from a genomic library of maize inbred line B73. Genomic libraries of other maize cultivars or plant species, as discussed above, are also suitable sources for isolating these genes. A preferred means for isolating clones from a genomic library is PCR amplification using genomic templates and *Sbe1* or *Ae*-specific primers derived from exons of SEQ ID NO:1 or SEQ ID NO:2.

In accordance with the present invention, genomic clones having the appropriate level sequence homology with the coding regions of SEQ ID NO:1 or 2 may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide.

- 22 -

Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 2X SSC and 0.1% SDS; (4) 2 hours at 45-55°C in 2X SSC and 0.1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology (Sambrook et al., 1989):

$$T_m = 81.5^{\circ}\text{C} + 16.6\text{Log} [\text{Na}^+] + 0.41(\% \text{G+C}) - 0.63 (\% \text{formamide}) - 600/\text{\#bp in duplex}$$

As an illustration of the above formula, using $[\text{Na}^+] = [0.368]$ and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C. Such a sequence would be considered substantially homologous to the sequences of the present invention. In a preferred embodiment, the hybridization is at 37°C and the final wash is at 42°C, in a more preferred embodiment the hybridization is at 42°C and the final wash is at 50°C, and in a most preferred embodiment the hybridization is at 42°C and final wash is at 65°C, with the above hybridization and wash solutions. Conditions of high stringency include hybridization at 42°C in the above hybridization solution and a final wash at 65°C in 0.1X SSC and 0.1% SDS for 10 minutes.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a

- 23 -

preferred embodiment, clones are maintained in plasmid cloning/expression vector, such as pBluescript SK- (Stratagene, La Jolla CA) which is propagated in a suitable *E. coli* host cell.

5 The expression-controlling elements of the invention, as mentioned above, include transcriptional control elements such as promoters and enhancers that occur in flanking regions, in introns and sometimes in the coding region of the gene, as well as translational
10 control elements (5' and 3') that affect the translation efficiency and/or stability of the mRNA. These elements are contemplated for use in a variety of applications in accordance with the present invention.

A preferred use of the expression-controlling
15 elements is to drive transcription and translation of native and heterologous genes in transgenic plants, for the purpose of increasing the production of the endogenous gene products, or for producing new gene products in the tissue- and developmentally-specific
20 manner governed by the expression-controlling sequences. Examples of useful gene products that can be expressed in seeds and other starch storage locations in plants include, but are not limited to: (1) gene products conferring herbicide tolerance; (2) gene products
25 involved in starch synthesis, such as starch synthases or ADP-glucose pyrophosphorylase; (3) gene products involved in fatty acid biosynthesis, such as fatty acid desaturases; (4) gene products which are seed storage proteins, such as zeins; (5) gene products conferring
30 resistance to insect pests, such as the crystal BT toxin; (6) gene products conferring resistance to microbial plant pathogens, including, e.g., viral coat proteins for virus resistance, or fungal cell wall lytic enzymes or

- 24 -

phytoalexins for fungal resistance; and (7) gene products that comprise or produce pharmaceutical or biological agents, such as antibiotics, secondary metabolites, antibodies, peptides or vaccines.

5 To use the expression-controlling sequences for introducing native or foreign genes into plants, the selected codings sequences and expression-controlling sequences are operably linked to one another, then inserted into vectors (or used directly, in some
10 transformation protocols) for transforming plant cells. Methods for manipulating DNA sequences to accomplish this are well known in the art, as exemplified by Sambrook et al. and Ausubel et al. The design of such DNA constructs will depend on the expression functions desired for the
15 gene of interest. As one example, all of the 5' and 3' expression-controlling sequences, as well as the +28 - +228 region, of the maize *Sbe1* gene may be operably linked to a coding sequence, such as a cDNA encoding a zein protein. A similar construct can be made using all
20 of the 5' and 3' expression controlling elements of the *Ae* gene. In transgenic plants containing such chimeric genes, the zein protein would be produced in the same tissues under the same physiological conditions as the *SBEI* or *SBEIb* proteins are produced. In another
25 embodiment, if it is desired to produce a sugar-inducible gene, the segments of the 5' region of *Sbe1* controlling sugar regulation (-314 to -145) could be utilized independently.

 In addition, certain of the expression
30 regulatory elements (e.g., the -314 to -145 element or the +28 to +228 element of *Sbe1*) may be combined with constitutive or inducible promoters from other sources. Examples of constitutive promoters suitable for this

- 25 -

purpose are the rice actin promoter, the maize ubiquitin promoter and the Cauliflower Mosaic Virus (CaMV) 35S promoter. An example of a suitable inducible promoter is the tetracycline repressor/operator controlled promoter.

5 Other examples of ways to combine the various 5', 3' and internal expression-controlling elements of the *SbeI* and *Ae* genes will be apparent to the person of skill in the art.

Transgenic plants expressing the native or
10 heterologous chimeric genes described above can be generated using standard plant transformation methods known to those skilled in the art. These include, but are not limited to, *Agrobacterium* vectors, PEG treatment of protoplasts, biolistic DNA delivery, UV laser
15 microbeam, gemini virus vectors, calcium phosphate treatment of protoplasts, electroporation of isolated protoplasts, agitation of cell suspensions with microbeads coated with the transforming DNA, direct DNA uptake, liposome-mediated DNA uptake, and the like. Such
20 methods have been published in the art. See, e.g., Methods for Plant Molecular Biology (Weissbach & Weissbach, eds., 1988); Methods in Plant Molecular Biology (Schuler & Zielinski, eds., 1989); Plant Molecular Biology Manual (Gelvin, Schilperoort, Verma, eds., 1993); and Methods in Plant Molecular Biology - A Laboratory Manual (Maliga, Klessig, Cashmore, Gruissem & Varner, eds., 1994).

The method of transformation depends upon the plant to be transformed. The biolistic DNA delivery
30 method is useful for nuclear transformation, especially of monocotyledenous plants. Methods for performing transformation using particle bombardment are described in detail in the Examples.

In another embodiment of the invention,

- 26 -

Agrobacterium vectors are used to advantage for efficient transformation of plant nuclei. In this embodiment, *Agrobacterium* "superbinary" vectors have been used successfully for the transformation of maize (Ishida et al., Nature Biotechnology 14: 745-750, 1996) and rice (Hiei et al., Plant Journal 6: 271-282, 1994).

Using an *Agrobacterium* superbinary vector system for transformation, the selected coding sequence under control of the expression-controlling elements of the invention, as described above, is linked to a nuclear drug resistance marker, such as kanamycin resistance or phosphothricin herbicide resistance. *Agrobacterium*-mediated transformation of plant nuclei is accomplished according to the following procedure:

- (1) the gene is inserted into the selected *Agrobacterium* binary vector;
- (2) transformation is accomplished by co-cultivation of plant tissue (e.g., leaf discs) with a suspension of recombinant *Agrobacterium*, followed by incubation (e.g., two days) on growth medium in the absence of the drug used as the selective medium;
- (3) plant tissue is then transferred onto the selective medium to identify transformed tissue; and
- (4) identified transformants are regenerated to intact plants.

Regardless of the transformation system used, it should be recognized that the amount of expression, as well as the tissue specificity of expression of the gene in transformed plants can vary depending on the position of their insertion into the nuclear genome. Such position effects are well known in the art. For this reason, several nuclear transformants should be regenerated and tested for expression of the transgene.

- 27 -

Another use of the expression-controlling elements is to inhibit or prevent transcription and translation of native *Sbe1* or *Ae* genes in plants. For this use, the elements are used as antisense molecules to
5 block gene expression at critical points. Antisense molecules can be transiently expressed in plants by introduction as single-stranded molecules (DNA or RNA), according to known methods. Alternatively, vectors encoding the antisense molecules can be introduced into
10 plants and the antisense molecules thereafter produced by transcription of the encoding sequences on the vectors, using the above-described methods. In another embodiment, overexpression of *Sbe1* or *Ae* is induced to generate a co-suppression effect. This excess expression
15 serves to promote down-regulation of both endogenous and exogenous *Sbe1* or *Ae* genes.

The following specific examples are provided to illustrate embodiments of the invention. They are not
20 intended to limit the scope of the invention in any way.

EXAMPLE 1

25 Genomic Organization and Promoter Activity of Maize Starch Branching Enzyme I Gene

This example describes the isolation of a full-length maize genomic DNA fragment containing the entire
30 *Sbe1* gene. Structural and functional analysis of this gene revealed a complete genomic organization and demonstrated the transcriptional activity of its promoter region.

- 28 -

Materials and methods

PCR amplification. Maize (*Zea mays* L., inbred W64A) genomic DNA prepared from 22-DAP kernels using the method of Rogers and Bendich (1985) was amplified in a 50 μ L reaction mixture containing 1 μ g of target DNA, 1 μ M of each primer, 200 μ M of each dNTP, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin (w/v), and 2.5 U of *Taq* DNA polymerase (Boehringer Mannheim, Indianapolis, IN). The mixture without dNTP was overlaid with 50 μ L of mineral oil (Sigma, St. Louis, MO) and incubated for 5 min at 94°C. The dNTPs were added to the mixture at the end of the incubation and then the mixture was cycled 35 times in a thermal cycler (ERICOMP, San Diego, CA) as follows: 94°C for 30 s; 55°C for 1 min; and 72°C for 2 min; with a final 72°C extension of 7 min. The primers were designed according to the published sequence data of the maize partial *Sbe1* cDNA (Baba et al., Biochem. Biophys. Res. Commun. 181: 87-94, 1991). The 5' primer, 5'-GACTGAATTCCTGCGCAGGAGGCAGAGCTT-3' (SEQ ID NO:3), and the 3' primer, 5'-GATCGAATTCCCATAGATACG TGGAGCAGCA-3' (SEQ ID NO:4), are homologous and complementary to DNA sequences of the maize *Sbe1* cDNA from 438 bp to 457 bp and 745 bp to 764 bp, respectively. Each primer contains an *EcoR* I restriction enzyme site and 4 extra nucleotides (underlined) at their 5' ends for convenience of subsequent cloning of the PCR product. After amplification, 15 μ L of the reaction sample was run on an agarose gel (1.5 %, w/v) in 1X TAE buffer containing 0.04 M Tris-acetate and 1 mM EDTA. A single PCR product of about 0.5 kb was detected on an ethidium bromide-stained agarose gel, digested with *EcoR* I restriction enzyme and cloned into the corresponding site of pBluescript SK⁻ (Stratagene, La Jolla, CA) creating

- 29 -

plasmid pB1.

Maize genomic library screening and DNA

sequencing. An EMBL-3 genomic library (Clontech, Palo Alto, CA) prepared from maize seedlings (2-leaf stage, B73) was screened essentially according to Sambrook et al. (1989). Approximately 3×10^5 pfu were transferred onto nylon membranes (Hybond-N+, Amersham, UK) and hybridized with the ^{32}P -labeled *SbeI* genomic PCR product excised from pB1. Hybridization was performed at 55°C for 20 h in 0.5 M Na_2HPO_4 (pH 7.2) and 7% SDS with gentle agitation at 40 cycles per minute on a rotary shaker. Following the hybridization, filters were washed twice in 5% SEN (5% SDS (w/v), 1 mM EDTA, 0.04 M Na_2HPO_4 , pH 7.2) and once in 1% SEN (1% SDS, 1 mM EDTA, 0.04 M Na_2HPO_4 , pH 7.2) for 15 min at 65°C . Plaques strongly hybridizing to the probe were selected and purified through three rounds of screening. Phage DNAs were isolated from the positive plaques according to Chisholm's method (1988) and were digested with *Pst* I to release the inserts from the EMBL-3 vector. The restriction fragments were separated on a 0.8% agarose gel and blotted onto nylon membranes. The blots were probed with ^{32}P -labeled full-length *SbeI* cDNA (Fisher et al., 1995, *supra*) as described above and hybridizing DNA fragments were identified and subcloned into pBluescript SK-. DNA sequences were determined by the dideoxynucleotide chain termination method with Sequenase Version 2.0 (United States Biochemical Co., Cleveland, OH). Sequence analyses were performed using programs from DNASTAR Inc. (Madison, WI).

Primer extension analysis. To locate the transcription initiation site of the *SbeI* gene, an oligonucleotide, 5'-GGCGACACGAGGCACAGCAT-3' (SEQ ID NO:5), which is complementary to the sense strand

- 30 -

sequence of the *Sbe1* cDNA from +1 to +20 relative to the translation start site (ATG) was radiolabeled at its 5' terminus with T4 polynucleotide kinase and $\gamma^{32}\text{P}$ -ATP. Approximately 10^5 cpm of the labeled primer was

5 hybridized at 35°C with 10 μg of total RNA, which was isolated from 30-DAP maize kernels (B73) according to the protocol of Vries et al. (Plant. Mol. Biol. Manual B6: 1-13, 1988). After hybridization for 8 h, complementary

10 DNA was synthesized from the annealed primer by the addition of reverse transcriptase and dNTP. Following the addition of EDTA and RNAase A into the reaction, the nucleic acid was precipitated with ethanol. The reaction

15 products were resuspended in sequencing gel loading buffer, denatured at 95°C, electrophoresed through a 5% polyacrylamide sequencing gel (w/v), and visualized by autoradiography. In order to provide size markers, part of the *Sbe1* gene was sequenced with the same primer used in the primer extension experiment.

3'-Rapid amplification of cDNA ends (3'-RACE).

20 To isolate the 3' end of the *Sbe1* transcript, the 3'-RACE method was used (Frohman et al., 1988). First-strand cDNA synthesis reaction was performed as follows: 14 μL of a mixture containing 5 μg of total RNA from 30-DAP maize kernels (B73) and 50 pmol of a 39-bp

25 oligonucleotide with 17 dT residues and an adaptor sequence, 5'-GGTCGACTCGAGTCGACATCGATTTTTTTTTTTTTTTT-3' (SEQ ID NO:6), was heated at 70°C for 10 min and quickly chilled on ice. To the chilled sample, 2 μL of 10X synthesis buffer (200 mM Tris-HCl, pH 8.4, 500 mM KCl,

30 25 mM MgCl_2 , and 1 mg/mL BSA), 1 μL of 10 mM dNTP mix, 2 μL of 0.1 M DTT and 1 μL (200U) of SUPERScript reverse transcriptase (GIBCO BRL, Grand Island, NY) were added and incubated at room temperature for 10 min. The

- 31 -

reaction mixture was then placed in a 42°C water bath for 50 min and transferred to a 90°C water bath. After 5 min incubation, 1 μ L of RNase H (2U/ μ L) was added and incubation at 37°C was continued for 20 min. Next, the first-strand cDNA obtained was amplified directly by the PCR method using a gene-specific primer (5'-GACTGAGCTCA TACCAAATGAAGCCAGGAG-3' (SEQ ID NO:7)), which is a homologous to sequence of the *Sbe1* gene from +5382 to +5401, and an adaptor primer (5'-GGTCGACTCGAGTCGACATCGA-3' (SEQ ID NO:8)). After amplification, a single DNA band detected on an agarose gel (1.5%, w/v) was isolated and digested with *Sac* I and *Xho* I (underlined within the primers). The resulting fragment, approximately 350 bp in length, was cloned into a pBlusScript SK⁻ (pB13R) and sequenced.

Genomic DNA blot analysis. Maize genomic DNA was prepared from 7-day-old etiolated seedlings (inbred B73) according to the method described by Junghans and Metzlauff (Junghans and Metzlauff, *Biotechniques*, 8: 176, 1990). 10 μ g of genomic DNA was digested with restriction enzymes *Bam*H I, *Eco*R I, *Bgl* II, and *Hind* III, separated on 0.8% agarose gels, and transferred onto nylon membranes (Hybond-N, Amersham, UK) in 20 X SSC containing 3 M NaCl and 0.3 M sodium citrate (pH 7.0) according to Sambrook et al. (1989). The DNA was crosslinked to the membrane by 3.5 min of UV irradiation on a transilluminator (312 nm). The genomic blots were prehybridized at 65°C for 1 h in 0.5 M Na₂HPO₄ (pH 7.2), 7% SDS, and 100 μ g/mL denatured salmon sperm DNA. Using the random primed DNA labeling kit (Boehringer Mannheim, Indianapolis, IN), 25 ng of a full-length *Sbe1* cDNA (Fisher et al., 1995, *supra*) was labeled with [α -³²P]-dCTP. This labeled probe was added to the

- 32 -

prehybridization solution and incubated at 65°C for 18 h. Blots were washed twice in 5% SEN and once in 1% SEN for 15 min at 65°C and were exposed to Kodak X-AR film at - 80°C for 48 h using two intensifying screens.

5 **Construction of *SbeI* promoter-*UidA* expression plasmids.** To make a transcriptional chimeric construct consisting of the *SbeI* promoter (-2190 to +27) fused to a β -glucuronidase (GUS) reporter gene, *UidA*, a *BamH* I restriction enzyme site was created just before the
10 translation initiation site of the *SbeI* gene as follows: the DNA sequence between -253 and +27 of the *SbeI* gene was amplified via polymerase chain reaction using *Pfu* DNA polymerase (Stratagene, La Jolla, CA) to enhance the fidelity of PCR amplification. The 5' primer, 5'-
15 CCAGCTCCACGGTTGTT CGTGT-3' (SEQ ID NO:9), is homologous to sequence of the *SbeI* gene from -253 to -232. An *Apa* I restriction enzyme site (GGGCCC) is located immediately downstream of the 5'-primer-binding region of the *SbeI* promoter, -203 to -198. The 3' primer, 5'-CGATGGATCC
20 TGTGACGGCGTGTGAGT CCC-3' (SEQ ID NO:10), consists of a DNA sequence complementary to that of the *SbeI* gene from +8 to +27 and a *BamH* I restriction enzyme site (GGATCC) flanked with four random nucleotides (underlined). The PCR product was digested with *Apa* I and *BamH* I, and the
25 resulting 236-bp fragment was cloned into pBluescript SK⁻ and sequenced in order to verify that no misincorporation had occurred in the DNA sequence during the PCR amplification. To incorporate this mutation into the context of a longer promoter fragment, the *SbeI* genomic
30 clone (λ 5-1-1) was digested with *Sal* I to isolate the approximately 3.0-kb fragment, which was then blunt-ended with Klenow and ligated with *Pst* I linkers (New England Biolabs). After complete digestion with *Pst* I and *Apa* I,

- 33 -

the resulting 2-kb *SbeI* promoter fragment was ligated to the 236-bp *Apa I*-*BamH I* fragment and cloned into plasmid pBI221 cut with *Pst I* and *BamH I*, thereby creating plasmid pKG101.

5 **Transient gene expression assay.** Suspension culture cells of maize endosperm (inbred A636), kindly provided by J. L. Anthony (DEKALB Genetics Corporation, Mystic, CT) were grown in 250-mL large-mouth Erlenmeyer flasks containing 80 mL of Murashige and Skoog basal salt
10 medium (Murashige and Skoog, 1962) supplemented with 0.4 mg/L thiamine, 2 g/L asparagine, and 30 g/L sucrose (Shannon and Liu, 1977). The culture was maintained in the dark at 29°C on a rotary shaker (120 rpm) and was subcultured every 7 days by transferring a portion of the
15 cell suspension into fresh medium. For particle bombardment, the growing cells (3 days after subculture) were evenly distributed over the surface of three layers of filter paper (Whatmann #4, 55 mm in diameter) moistened with 3 mL of the liquid medium and positioned
20 in the middle of a 10-cm petri dish. Three milligrams of 1.6- μ m gold particles were coated with 10 μ g of pKG101 according to the method described by Xu et al. (Xu et al., Plant Mol. Biol., 31: 1117-1127, 1996) and introduced into the cells using a Bio-Rad PDS-1000/He
25 Biolistic Particle Delivery system. Bombardments were performed at 650 psi under a vacuum of 26 inches of Hg with a distance of 10 cm between the cells and the microprojectile launch site of the particle gun. Following the bombardments, the petri dishes were sealed
30 with parafilm and then incubated in the dark at 25°C for 24 h.

Histochemical GUS staining. Histochemical assays of GUS activity were performed according to the

- 34 -

method described by Jefferson et al. (Jefferson et al., Plant Mol. Biol. Rep., 5: 387-405, 1987) with minor modifications. Briefly, after 24 h incubation, cells on the filter paper were transferred into petri dishes containing 2 mL of GUS staining solution (50 mM sodium phosphate buffer, pH 8.0, 0.05 mM potassium ferricyanide, 0.05 mM potassium ferrocyanide, 0.2% Triton X-100 (w/v), 20% methanol (v/v), and 1 mg/mL 5-bromo-4-chloro-3-indolyl- β -D-glucuronide). The petri dishes were then sealed and incubated at 37°C overnight. The cells were examined and photographed under a Nikon SMZ-U dissecting microscope.

Results

PCR amplification of maize genomic DNA. In order to obtain a DNA probe for genomic library screening, maize genomic DNA prepared from 22-DAP kernels (W64A) was amplified by polymerase chain reaction (PCR) using upper and lower primers designed to anneal the *Sbe1* cDNA (Baba et al., Biochem. Biophys. Res. Commun., 181: 87-94, 1991) from 438 to 457 and 745 to 764, respectively. A single amplified DNA band, approximately 450 bp in length, was observed on an ethidium bromide-stained agarose gel. The PCR product containing *EcoR* I sites at both 5'- and 3'-ends was digested with *EcoR* I and the resulting fragment was cloned and sequenced.

Alignment of sequences between the maize PCR genomic fragment and the published *Sbe1* cDNA (Baba et al., 1991, *supra*) shows that the PCR fragment contained the predicted *Sbe1* cDNA sequence with only 3 nucleotide differences and a 103-bp intron. The differences could be explained by cultivar polymorphisms or misincorporation of bases during PCR amplification by Tag

- 35 -

DNA polymerase which does not have proofreading activity. This PCR fragment was ³²P-labeled and used as a hybridization probe for screening a maize genomic library to isolate *Sbe1* genomic clones.

5 **Isolation and analysis of a maize *Sbe1* genomic**

clone. In a screen of approximately 3×10^5 plaque-forming units from a genomic library prepared from maize seedlings (inbred B73), 8 positive lambda clones were isolated which strongly hybridized to the probe.

10 Restriction mapping and partial DNA sequencing of these clones indicated they all probably originated from the same genetic locus. A full-length clone (λ 5-1-1), containing the entire coding region of the *Sbe1* gene, as well as 5'- and 3'- flanking sequences, was selected for
15 further analyses. Figure 1A shows a restriction map of the genomic clone. The 3.0-kb *Sal* I fragment and the 6.2-kb *Pst* I fragment from the clone were subcloned into pBluescript SK⁻ producing plasmids pBI5-1 and pBI5-2, and their nucleotide sequences were completely determined.

20 A consensus TATA-box as well as a G-box containing a perfect palindromic sequence known as a G-box, CCACGTGG, were found in the 5' flanking region of the gene (Fig. 2). Primer extension analysis displayed a single extended product of 44 nucleotides, which co-
25 migrates with an A residue in the sequencing ladder generated with the *Sbe1* 5'-flanking region. This indicates that consistent with many eukaryotic genes, the transcription initiates at a position which is located 25 bp downstream from the putative TATA box, suggesting that
30 the TATA box may be a functional element of the *Sbe1* promoter.

To determine the polyadenylation site of the *Sbe1* gene, 3' RACE was conducted. It demonstrated that a

- 36 -

poly (A) tail occurs at an adenine nucleotide located 29 bp downstream from a putative polyadenylation signal (AATAAA) in the *Sbel* gene. Along with the primer extension result, this indicated that the transcribed region in the *Sbel* gene is 5,690 bp in length.

Structure of the *Sbel* genomic clone. Alignment of the genomic sequence with the published maize *Sbel* cDNA sequence (Fisher et al., 1995, *supra*) revealed that the gene is composed of 14 exons and 13 introns distributed over 5.7 Kb. Figure 1B summarizes the organization of the maize *Sbel* gene. The cDNA sequence is identical to the corresponding genomic sequence except for a 1-bp mismatch in exon 14. Table 1 shows the sequences around the exon/intron junctions and a list of putative branch point consensus sequences, which were derived as described by Brown (Brown, J.W.S., Nucl. Acids Res., 14: 9549-9559, 1986).

Table 1. List of introns and sequences of exon/intron borders in the *Sbel* gene

Intron number	Exon /	Putative intron branch point*	/Exon	Intron size(bp)	GC content (%)
1	TCGCG	<u>GTAAG</u> <u>CTGAT</u> ·17 ^y · <u>CGCAG</u>	GGTGG	82	46.3
25 2	GGAAG	<u>GTAGA</u> <u>CTGAA</u> ·24· <u>GGAAG</u>	GTCAA	377	36.9
3	TAAAG	<u>GTTAG</u> <u>ATCAT</u> ·30· <u>TTCAG</u>	GCTAT	120	34.2
4	GCGCA	<u>GTAAG</u> <u>TTGAC</u> ·25· <u>TGTAG</u>	GGAGG	319	34.2
5	GAAAG	<u>GTCTC</u> <u>ATGAC</u> ·33· <u>TGCAG</u>	GTACA	103	35.9
6	ATCAG	<u>GTACC</u> <u>ATGAC</u> ·43· <u>TTCAG</u>	TCTAT	262	32.1
30 7	AAAAG	<u>GTTCC</u> <u>CTCAA</u> ·33· <u>TCCAG</u>	ATGAT	97	39.2
8	ATGAG	<u>GTGAA</u> <u>TTGAT</u> ·17· <u>TCTAG</u>	TTTGG	488	35.2
9	ACAAG	<u>GTTAT</u> <u>CTAAC</u> ·37· <u>AACAG</u>	TACAT	565	37.3
10	AAAAG	<u>GTAAG</u> <u>GTCAG</u> ·25· <u>TTCAG</u>	GTTAT	160	39.4
11	GAGGG	<u>GTAAG</u> <u>CTTAC</u> ·31· <u>TGCAG</u>	CTACA	107	41.1
35 12	GAAGA	<u>GTAAG</u> <u>CTCAT</u> ·16· <u>CGCAG</u>	GTTGG	140	49.3
13	GTGTG	<u>GTAAT</u> <u>CTGAC</u> ·18· <u>GCCAG</u>	GCTTA	73	49.3

* Consensus sequences between introns are underlined.

^y Numbers indicate number of nucleotides between adjacent sequences.

- 37 -

The introns, relatively AT-rich (61%) compared to the exons (52%), vary in length from 73 bp to 565 bp, and all of which have the conserved sequences at their 5' and 3' ends, following the 'GT . . AG' rule of plant introns (Brown, 1986, *supra*). Exon 1, containing 27 bp of 5'-untranslated DNA sequence, and exon 2 occur in the transit peptide region which may be essential for transporting the gene product into the amyloplast. Exon 14 contains the translation stop codon (TGA) and 3'-untranslated region, as well as the putative polyadenylation signal (AATAAA). The exons vary in length from 63 to 907 bp.

Comparison of the maize and rice *Sbe1* genomic DNA sequences (Kawasaki et al., Mol. Gen. Genet., 237: 10-16, 1993b) revealed two large highly conserved regions in the 5'-flanking sequences (Fig. 2). One region, 161 bp in length, was present between -2190 and -1890 in the maize *Sbe1* 5'-flanking sequence, and has 82% similarity with the corresponding rice region. The other, 342 bp in length, is located from -1804 to -1611, and shows 85% similarity with the corresponding rice region. These sequences conserved between the species suggest that the regions may play an important role in gene expression. Interestingly, a part of the latter region from -1804 to -1611 (194 bp in length) shares 83% similarity with the 5'-untranslated sequence of a Ca²⁺-dependent protein kinase gene in rice (Kawasaki et al., Gene., 129: 183-189, 1993a), suggesting that a maize version of protein kinase gene may be located immediately upstream of the *Sbe1* gene. In the rice genome, the two genes are separated by approximately 1.4 kb, and transcribed divergently from each other.

5'-flanking sequences downstream from the two

- 38 -

conserved regions did not show sequence similarity between the two genes (less than 25%) except for a G-box motif and several small segments adjacent to the G-box. This may indicate that the G-box plays a role in regulation of *Sbel* gene expression. The G-box is found in other plant genes which respond to diverse environmental or physiological stimuli and are often associated with additional regions which possibly act as coupling elements determining signal response specificity.

Another notable feature derived from the sequence comparison is that the maize and rice *Sbel* genomic structures are quite similar. Both genes consist of 14 exons and 13 introns, the positions of which are conserved between the two species. The sizes of exons (exon 3 to exon 13) constituting most of the mature proteins are identical except exon 5, in which the rice *Sbel* gene has one more codon compared to the maize gene. They share more than 86% and almost 90% similarity in nucleotide and amino acid sequences, respectively. However, exons not encoding the mature proteins (exon 1, 2 and most of exon 14) do not display any significant sequence similarity and vary in size. The carboxyl-terminal 64 (67 in rice) amino acids encoded by exon 14 in the maize gene is the only region which is not conserved in the mature protein. Unlike the exons, homology was not found in any of the introns other than the splice junction sequences. The large intron (2212 bp) present in the rice *Sbel* gene is not found in the maize gene.

Genomic Southern blot analysis. To determine the number of *Sbel* genes in the maize genome, Southern blot analysis was performed. When blots were hybridized

- 39 -

with the full-length maize *Sbe1* cDNA (Fisher et al., 1995, *supra*) under high-stringency conditions, at least three bands were detected in each lane. Comparison of the hybridization patterns with the restriction map of the *Sbe1* genomic clone (Fig. 1A) revealed that not all
5 the bands in the Southern blot corresponded to the genomic map, suggesting more than one *Sbe1* gene is present in the maize genome.

To confirm this, a 0.6-kb genomic DNA probe
10 which did not have any restriction enzyme sites used in the genomic blot was prepared from the genomic clone λ 5-1-1 by *Bam*H I-*Hind* III digestion. The genomic probe will produce only one hybridizing band in every lane if there is a single copy of the *Sbe1* gene in the maize genome.
15 However, the genomic probe (probe 2) containing the central region of the *Sbe1* cDNA detected in each lane one or two additional bands apart from the bands predicted by the genomic map. This indicates that along with the isolated *Sbe1* gene, another *Sbe1* gene or a gene very
20 closely related to *Sbe1* exists in the maize genome. These will be referred to as *Sbela* (identified) and *Sbelb* (unidentified) to distinguish them when appropriate. Interestingly, when a 1.7-kb *Bam*H I-*Pst* I genomic fragment consisting of the *Sbela* promoter and transit
25 peptide-coding region was used as a probe (probe 1), only the bands predicted from the identified *Sbela* genomic DNA sequence were detected. Taken together, these results suggest that although two *Sbe1* genes are present in the maize genome, their 5' flanking sequences and the 5'-end
30 of the coding regions (at least in DNA level) are quite divergent from each other.

Genetic mapping of *Sbe* genes. In order to map *Sbe1* loci onto the existing framework map (Gardiner et

- 40 -

al., Genetics, 134: 917-930, 1993), the full length *Sbe1* cDNA was used to probe a population of 54 immortalized F₂ individuals from the cross Tx303 X CO159 as part of the mapping efforts of the University of Missouri-Columbia
5 Maize RFLP Laboratory. Two loci were identified on chromosome 6, bin 6.01 and chromosome 10, 10.04. This supports the conclusion made above based on genomic Southern analysis that two *Sbe1* genes exist in the maize genome. However, at present the data do not allow us to
10 distinguish which gene is located on which chromosome.

Promoter activity. A transient expression system was used to test whether the 5'-flanking sequence of the cloned *Sbe1* gene (λ 5-1-1) is sufficient to support transcriptional activity. A chimeric gene
15 containing a 2.2-kb 5' *Sbe1* fragment (-2191 to +27) fused to *UidA* reporter gene in pUC19 was first constructed and designated pKG101. The chimeric plasmid was introduced into maize endosperm suspension cells via particle bombardment. Iodine staining and northern blot analysis
20 showed that the maize endosperm suspension cells actually produce starch and the genes involved in starch biosynthesis are expressed. The bombarded cells were incubated for 24 h at 25°C in the dark and histochemical GUS assays were performed to visualize GUS expression.
25 Blue spots were observed, indicating that the 5'-flanking sequence has the ability to drive gene expression in the maize endosperm cells. Control cells bombarded with gold particles coated with promoterless *UidA* construct did not show any blue spots.

30

- 41 -

EXAMPLE 2**Molecular Cloning and Characterization
of the Amylose-Extender Gene Encoding
Starch Branching Enzyme IIB in Maize**

5

This example describes the complete genomic organization of the Ae gene and the promoter regions critical for its expression in maize endosperm cells.

10 **Materials and Methods****Maize Genomic Library Screening and DNA**

Sequencing. Using ^{32}P -labeled full-length Ae cDNA (Fisher et al., 1993, *supra*) as probe, an EMBL-3 genomic library (Clontech, Palo Alto, CA) prepared from maize seedlings (B73, 2 leaf stage) was screened essentially according to Sambrook et al. Approximately 3×10^5 plaque forming units were transferred onto nylon membranes (Hybond-N+, Amersham, UK). Hybridization was performed at 55°C for 20 h in a solution containing 0.5 M Na_2HPO_4 , pH 7.2, and 7% SDS with gentle agitation at 40 cycles per minute on a rotary shaker. Following the hybridization, filters were washed twice in 5% SEN (5% SDS (w/v), 1 mM EDTA, 0.04 M Na_2HPO_4 , pH 7.2) and once in 1% SEN (1% SDS, 1 mM EDTA, 0.04 M Na_2HPO_4 , pH 7.2) for 15 min at 65°C. Plaques strongly hybridizing to the probe were selected and purified through three rounds of screening. Phage DNAs were isolated from the positive plaques according to Chisholm's method and SalI-digested inserts were subcloned into pBluescript SK-. DNA sequences were determined by the dideoxynucleotide chain termination method with Sequenase Version 2.0 (United States Biochemical Co., Cleveland, OH). Sequence analyses were performed using programs from DNASTAR Inc. (Madison, WI).

- 42 -

Primer Extension Analysis. To locate the transcription initiation site of the Ae gene, an oligonucleotide, 5'-GATCGGATCGAACTGATCAG-3' (SEQ ID NO:11), which is complementary to the sense strand sequence of the Ae cDNA from -35 to -16 relative to the translation start site (ATG) was radiolabeled at its 5' terminus with T4 polynucleotide kinase and $\gamma^{32}\text{P}$ -ATP. Approximately 10^5 cpm of the labeled primer was hybridized at 35°C with 10 μg of total RNA, which was isolated from 30 DAP maize kernels (B73) according to the protocol of Vries et al. (In: Gelvin, SB, Schilperoot, RA (eds) Plant Molecular Biology Manual B6: 1-13, Kluwer Academic Publishers, Dordrecht, Netherlands, 1988). After hybridization for 8 h, complementary DNA was synthesized from the annealed primer by the addition of reverse transcriptase and dNTP. Following the addition of EDTA and RNAase A into the reaction, the nucleic acid was precipitated with ethanol. The reaction products were resuspended in sequencing gel loading buffer, denatured at 95°C, electrophoresed through a 5% polyacrylamide sequencing gel (w/v), and visualized by autoradiography. In order to provide size markers, part of the Ae gene was sequenced with the same primer used in the primer extension experiment.

Genomic Southern Blot Analysis. Maize genomic DNA was prepared from 7-day-old etiolated seedlings (inbred B73) according to the method described by Junghans et al (1990, *supra*). 10 μg of genomic DNA was digested with restriction enzymes, separated on 0.8% agarose gels, and transferred onto nylon membranes (Hybond-N, Amersham, UK) in 20 X SSC containing 3 M NaCl and 0.3 M sodium citrate, pH 7.0 according to Sambrook et al. DNA was crosslinked to the membrane by 3.5 min of UV

- 43 -

irradiation on a transilluminator (312 nm). Genomic blots were prehybridized at 65°C for 1 h in 0.5 M Na₂HPO₄, pH 7.2, 7% SDS, and 100 µg/mL denatured salmon sperm DNA. 25 ng of a full-length Ae cDNA (Fisher et al., 1993, supra) was labeled with [α-³²P]-dCTP using a random primed DNA labeling kit (Boehringer Mannheim, Indianapolis, IN). Labeled probe was added to the prehybridization solution and incubated at 65°C for 18 h. Blots were washed twice in 5% SEN and once in 1% SEN for 15 min at 65°C and were exposed to Kodak X-AR film at -80°C for 48 h using two intensifying screens.

Construction of Plasmids. For a transcriptional fusion of the Ae promoter to a luciferase (LUC) reporter gene, a BamHI restriction enzyme site was created just before the translation initiation site (ATG) of the Ae gene as follows. The DNA sequence between -14 and +100 of the Ae gene was first amplified via polymerase chain reaction. The primer (PII-2), 5'-CCTAATTGTAGCCCTGCAGTCA-3' (SEQ ID NO:12), is homologous to sequence of the Ae gene from -10 to +12. A PstI site (CTGCAG) is located immediately downstream of the primer binding region of the Ae promoter, +4 to +9. The 3' primer (PII-3), 5'-GACTGGATCCTCGCCTTCGCAGCCGGATCG-3' (SEQ ID NO:13), consists of a DNA sequence complementary to that of the Ae gene from +80 to +100 and a BamHI restriction enzyme site (GGATCC) flanked with four random nucleotides (underlined). The PCR product was digested with PstI and BamHI, and the resulting 100-bp fragment was ligated to the 2,977-bp SalI-PstI Ae promoter fragment and cloned into plasmid pLN cut with SalI and BamHI (promoterless LUC-NOS gene in pUC119), thereby creating plasmid pKL201. This construct as well as all the following constructs

- 44 -

were verified by DNA sequencing.

To construct a translational fusion of the Ae promoter containing the first exon and intron to a LUC reporter plasmid, the Ae genomic clone, 3-2-1, was
5 digested with XhoI and the resulting 866-bp fragment was gel purified. The fragment was blunt ended by Klenow fill-in DNA synthesis and ligated with BamHI linkers (CGGGATCCCG). After complete digestion with PstI and BamHI, the 325-bp DNA fragment was isolated and used to
10 replace the 100-bp PstI-BamHI region in pKL201. This construct was designated pKLN201.

A series of 5' deletion mutants were derived from pKL201 using available restriction enzyme sites and PCR techniques. To create pKL202, pKL203, pKL205 and
15 pKL206, pKL201 was first digested with AccI (-1719 to -1714), SpeI (-1128 to -1123), XhoI (-537 to -532) and ApaI (-348 to -343), respectively. Then, each linearized plasmid was separately gel purified and blunt ended by Klenow fragment. After SalI linker ligation,
20 the modified plasmids were digested with SalI, and the larger DNA fragments from each reaction were isolated and self-ligated to produce the relevant plasmids carrying different deletion end points.

For a construction of pKL204, the Ae promoter
25 region between -755 and -500 was amplified with two primers. The 5' primer containing a SalI (GTCGAC) flanked with four extra nucleotides (underlined), 5'-GAAAGTCGACGAAGAGAGAATGAAAGCGAA-3' (SEQ ID NO:14), and the 3' primer, 5'-GCGCGGGTCCGTCGTGCCTTTT-3' (SEQ ID NO:15),
30 were designed to anneal to DNA sequences of the Ae gene from -755 to -736 and from -521 to 500, respectively. The amplified 266-bp product was digested with SalI and XhoI which was located 10 bp upstream of the 3' primer

- 45 -

binding region, and the resulting 230-bp fragment was used to substitute the 2.4-kb SalI-XhoI fragment in pKL201.

To make pKL207, the Ae promoter region between
5 -160 and +100 was first amplified with 5' primer, 5'-
GATAGTCGACCGACGCGCAACGGCCTGCCT-3' (SEQ ID NO:16), and 3'
primer, 5'-GACTGGATCCTCGCCTTCGCAGCCGGATCG-3' (SEQ ID
NO:17), which contain SalI and BamHI sites, respectively,
along with four arbitrary extra bases (underlined).
10 Next, the PCR product was digested with SalI and BamHI
and the resulting 267-bp DNA fragment was then used to
replace the 3.1-kb SalI-BamHI fragment in pKL201. The
same 3' primer and method were used for a construction of
pKL208 except for the 5' primer, 5'-
15 GATCGTCGACCGCTCGTCTCCGTCCTATAT-3' (SEQ ID NO:18),
homologous to the DNA sequence of the Ae gene from -49 to
-32.

Particle Bombardment. Suspension culture cells
of maize (inbred A636) endosperm provided by J. L.
20 Anthony (DEKALB Genetics Corporation, Mystic, CT) were
grown in 250-ml large-mouth Erlenmeyer flasks containing
80 ml of Murashige and Skoog basal salt medium
supplemented with 0.4 mg/l thiamine 2 g/l asparagine and
30 g/l sucrose. The culture was maintained in the dark
25 at 29°C on a rotary shaker (120 rpm) and was subcultured
every 7 days by transferring a spoonful of the cell
suspension into 80 ml of fresh medium.

For particle bombardment, about 600 mg (fresh
weight) of actively growing cells 3 days after subculture
30 was evenly distributed over the surface of filter paper
(Whatmann #4, 55 mm in diameter) by vacuum filtration of
8 ml of suspension culture. The filter paper bearing the
cells was then placed over three layers of filter paper

- 46 -

(Whatmann #4, 70 mm in diameter) moistened with 5 ml of the liquid medium containing 12% sucrose and positioned in the middle of a 10 cm petri dish.

60 mg of gold microcarriers (1.6 μ m particle size) were washed three times with 1 ml of 100% ethanol and twice with 1 ml of sterile deionized H₂O, resuspended in 1 ml of sterile deionized H₂O, and dispensed in 50- μ l aliquots (3 mg/50 μ l). The *Sbe1* Promoter-LUC constructs and a GUS reference plasmid (pBI 221; Clontech, Palo Alto, CA) were co-precipitated onto gold particles as follows: under continuous vortexing, the following were added in order to each 50- μ l aliquot of gold particles: 5 μ l of DNA (8 μ g of LUC reporter plasmid, 4 μ g of GUS reference plasmid), 50 μ l of 2.5 M CaCl₂, and 20 μ l of 0.1 M spermidine (free base, tissue culture grade). The gold particles coated with DNA were pelleted in an Eppendorf centrifuge at 10,000 rpm for 10 sec, rinsed with 250 μ l of 100% ethanol, and resuspended in 60 μ l of 100% ethanol. Immediately after sonication, 8 μ l of the DNA-coated gold particles were pipetted onto the center of macrocarriers (Bio-Rad, Hercules, CA) and dried in a low humidity environment.

A Bio-Rad PDS-1000/He Biolistic Particle Delivery system was used for particle bombardment. Bombardment parameters which were optimized include He pressure, gap distance (distance from power source to macroprojectile), and target distance (distance from microprojectile launch site to sample target). After optimization, all bombardments were performed in a dim room at 650 psi under a vacuum of 26 inches of Hg with a distance of 10 cm between the cells and the barrel of the particle gun. Following the bombardments, the petri dishes were sealed with Parafilm and then incubated in

- 47 -

the dark at 25°C for 24 hr.

GUS and LUC Assays. The bombarded cells were harvested from the plates by vacuum filtration, frozen in liquid nitrogen, and ground with a pestle and mortar to a fine powder. The powder was then transferred into a microfuge tube and extracted with cell culture lysis buffer containing 300 mM Tris-phosphate, pH 7.8, 2 mM DTT, 2 mM 1,2-diaminocyclohexane-N, N, N', N'-tetraacetic acid, 10% glycerol and 1% Triton X-100 (0.3 ml/g of tissue). Cell debris was pelleted in an Eppendorf centrifuge at 14,000 rpm for 10 min at 4°C and the supernatant was split into two aliquots for assays of GUS and LUC activity.

For fluorometric GUS assays, 30 μ l of the crude extract was incubated at 37°C with 2 mM 4-methylumbelliferyl glucuronide in 0.3 ml of GUS assay buffer (50 mM NaPO₄, pH 7, 10 mM EDTA, 0.1% Triton X-100, 0.1% Sarkosyl, 10 mM β -mercaptoethanol, 20% methanol). After 0, 1, and 2 hr of incubation, 0.1 ml aliquots were removed and added to 0.9 ml of 0.2 M Na₂CO₃ to terminate the reaction. A TKO 100 fluorometer (Hoeffer, San Francisco, CA) calibrated by setting a 100-nM MU to 1,000 fluorescence units was used to measure fluorescence of the product, 4-methyl umbelliferone (4-MU). For each sample, results of GUS assay were plotted in a graph of OD₄₀₅ (Y-axis) versus time in minutes and the GUS activity was expressed simply as the slope of the line. GUS activity from the maize endosperm suspension cells that had been bombarded with the naked gold particles (no DNA) was used as a control.

Using a luminometer (Monolight 1500; Analytical Luminescence Laboratory, San Diego, CA), luciferase activity was determined by measuring luminescence for 10

- 48 -

sec after mixing 20 μ l of cell extract with 100 μ l of luciferase assay reagent containing 20 mM tricine, pH 7.8, 1.07 mM $(\text{MgCO}_3)_4\text{Mg}(\text{OH})_2 \cdot 5\text{H}_2\text{O}$, 2.67 mM MgSO_4 , 0.1 mM EDTA, 33.3 mM DTT, 270 μ M coenzyme A, 470 μ M luciferin and 530 μ M ATP. LUC activity from the maize endosperm suspension cells that had been bombarded with the pLN (promoterless LUC plasmid) was used as a control. To correct differences in sample variability and transfection efficiency, the luciferase activity (in light units) was normalized with GUS activity, yielding the LUC/GUS ratio of each sample.

Results

Cloning and Characterization of the Ae Gene.

After screening approximately 3×10^5 plaque-forming units of a genomic library prepared from maize seedlings (inbred B73), 12 lambda clones that strongly hybridized to the full-length Ae cDNA probe were isolated. These clones were hybridized to 5' or 3' cDNA fragment probes revealing that none of them contained both ends of the Ae gene. Based on restriction endonuclease maps, two clones, λ 3-2-1 and λ 7-2-1, were selected, subcloned, and sequenced. DNA sequences of the two clones revealed that the λ 3-2-1 clone containing the 5'-end of the Ae gene had approximately 1.5-kb of overlap with the λ 7-2-1 clone containing the 3'-end of the gene. A complete restriction map of the Ae gene was constructed by combining the two overlapping genomic clones which encompasses the entire coding region of the gene as well as large regions of 5'- and 3'- flanking sequences (Fig. 3).

Genomic Organization of the Ae Gene. Primer extension analysis was conducted to determine the

- 49 -

transcription initiation site of the Ae gene. A single major reverse transcription product was observed which co-migrated with a G residue in the sequencing ladder, indicating the transcription initiates mainly at a position which is located 28 bp downstream from a putative TATA box. The transcription initiation site was numbered +1 in the sequence shown in Figure 4. In addition to the TATA box, a number of potential *cis*-regulatory elements were found in the 5'-flanking region of the Ae gene (Fig. 4). The proximal promoter region from -300 to -87 contains two MRE boxes (TGCRNC, R = purine, Y = pyrimidine), motifs essential for metal ion-dependent induction of both mouse and human metallothionein genes. In addition, this region contains four GC boxes (CCGCCC), sequences recognized by the mammalian transcription factor Sp1. In a region further upstream, sequences identical to a Hex (ACGTCA), a conserved element found in plant histone gene promoters; an I box (GATAGG), an element conserved in various RBCS genes encoding the small subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase; and a RY repeat (ATGCCATG), a distal regulatory element which comprises a portion of the 28-bp legumin box were present at positions -999, -647, and -1491, respectively.

Interestingly, DNA sequences from -505 to +463 are extremely high in G+C content (66.5%) and CpG dinucleotide frequency (10.1 per 100 bp) compared to an average of 40% G+C and 3.2 CpG per 100 bp for the rest of the genomic fragment. This region includes the proximal promoter region, two exons, and one intron. These characteristics are typical of CpG islands found in the mammalian genome, which are usually nonmethylated and flanked by methylated regions.

The genomic structure of the Ae gene was established by alignment with the published sequence of Ae cDNA. The transcribed region of the gene consists of 22 exons and 21 introns distributed over 16,914 bp in length. Figure 3 summarizes the organization of the Ae gene. The published cDNA sequence is identical to the corresponding genomic sequence except for three nucleotides present in exon 4 and exon 6. This may be due to the different genetic stocks used in the two studies. Table 2 shows the sequences around the exon/intron junctions and a list of putative branch point consensus sequences.

Table 2. List of introns and sequences of exon/intron borders in the Ae gene.

Intron number	Exon/	Putative intron branch point*	/Exon	Intron size(bp)	GC content (%)
1	ACTC	<u>GTAA</u> <u>GTGAA</u> ..24 ^y . <u>GCAG</u>	GGGG	106	51.9
2	GGAG	<u>GTTC</u> <u>CTGAA</u> ..29.. <u>CCAG</u>	GTAC	244	41.4
3	ATCG	<u>GTAT</u> <u>TTCAA</u> ..21.. <u>ACAG</u>	GTAC	1086	43.9
4	GCAG	<u>GTAT</u> <u>ATAAC</u> ..24.. <u>GTAG</u>	CGCG	76	25.0
5	ATTT	<u>GTAT</u> <u>TTCAG</u> ..25.. <u>TTAG</u>	TCTG	196	31.1
6	CAAA	<u>GTAT</u> <u>CTAAA</u> ..22.. <u>GCAG</u>	AATG	499	33.5
7	AAAG	<u>GTAG</u> <u>TTAAC</u> ..23.. <u>ATAG</u>	GTGA	81	40.7
8	AAGA	<u>GTCT</u> <u>TTAAG</u> ..21.. <u>GCAG</u>	GTAA	567	30.3
9	CCCG	<u>GTAT</u> <u>ATAAT</u> ..23.. <u>TTAG</u>	GAAC	774	32.0
10	TTGG	<u>GTAA</u> <u>TTAAT</u> ..21.. <u>GCAG</u>	ATAC	751	33.2
11	ATAG	<u>GTAA</u> <u>TTAAC</u> ..22.. <u>GCAG</u>	TCAT	4020	44.0
12	GGAA	<u>GTAC</u> <u>TTTAT</u> ..49.. <u>GCAG</u>	GTTT	86	34.9
13	ACAA	<u>GTAA</u> <u>CTAAA</u> ..19.. <u>TTAG</u>	GTAA	148	31.1
14	GTAA	<u>GTGC</u> <u>TTCAA</u> ..20.. <u>TCAG</u>	GTTA	3051	41.8
15	TCAA	<u>GTAA</u> <u>TTCAA</u> ..19.. <u>ACAG</u>	GCAA	872	31.8
16	CAAG	<u>GTTA</u> <u>ATGAG</u> ..25.. <u>GCAG</u>	GATA	457	32.6
17	CCTG	<u>GTGA</u> <u>GTCAT</u> ..21.. <u>GCAG</u>	AATG	144	30.6
18	CCTG	<u>GTAA</u> <u>ATTAT</u> ..28.. <u>TCAG</u>	GGTG	226	30.1
19	TGAA	<u>GTAT</u> <u>ATGAA</u> ..19.. <u>GCAG</u>	TTCA	266	31.2
20	TAAG	<u>GTAT</u> <u>CTGAC</u> ..21.. <u>CCAG</u>	GTGG	448	40.0
21	CGCC	<u>GTAA</u> <u>CTAAC</u> ..24.. <u>GCAG</u>	GACT	96	45.8

* Consensus sequences between introns are underlined.

^y Numbers indicate number of nucleotides between adjacent sequences.

- 51 -

The introns are relatively AT-rich (61%) compared to the exons (54%), and all have the conserved splice site sequences at their 5' and 3' ends, following the 'GT...AG' rule of plant introns. The introns vary in
5 length from 76 bp (intron 4) to 4,020 bp (intron 11), and the exons vary in length from 43 bp to 303 bp. Exon 1 contains 100 bp of 5' untranslated DNA sequence, and exon 22 contains the translation stop codon (TGA) and 3' untranslated region. Although the canonical
10 polyadenylation signal, AATAAA, was not found in the 3'-end of the gene, a similar sequence (AATTAAA) was observed 29 bp upstream of the polyadenylation site.

In addition, sequence analysis revealed that the 3'-flanking region of the Ae gene contains many
15 direct repeat sequences and has a high degree of similarity to the pollen retroelement maize-2 (PREM-2), a copia-type retroelement in maize which is expressed in a tissue-specific manner (Fig. 4). A 13-bp polypurine tract (AAAAAGGGGGAGA) which is present in PREM-2 and
20 necessary for retroelement replication was also found upstream of the region which is very similar to the 3' long terminal repeat (LTR) of the PREM-2. Thus, the 3'-flanking region is likely to possess part of a PREM-2-type retroelement.

25 **Genomic Southern Blot Analysis.** Southern blot analyses were performed to determine the number of genes in the maize genome that are similar to Ae. When blots were probed with the full-length Ae cDNA under high-stringency conditions, at least two strongly hybridizing
30 bands were observed in each lane. The band patterns agreed with the restriction map of the Ae genomic DNA, suggesting all the bands were derived from a single genetic locus. To confirm this, the blots were probed

- 52 -

with a small fragment of the Ae genomic DNA which does not have any restriction enzyme sites for BamHI, EcoRI, BglIII and HindIII. As expected, only a single band was observed in every lane, supporting the conclusion that a single copy of Ae is present in the maize genome.

Analysis of the 5'-flanking Region of the Ae Gene. To identify the 5'-flanking regions necessary for Ae gene expression, we utilized a transient expression assay system in maize endosperm suspension cells. Iodine staining and northern blot analysis showed that the maize endosperm suspension cells actually produce starch and the genes involved in starch biosynthesis are expressed. In an initial experiment, a transcriptional chimeric construct containing the Ae gene fragment between -2964 and +100 linked to a luciferase (LUC) reporter gene in pUC119 was created and called pKL201. Since there are several examples of plant genes which are regulated by the first exon and/or intron sequences, a translational fusion construct (pKLN201) containing the corresponding region of the Ae gene was also created to test its effect on gene expression. pKLN201 was created by including an additional Ae DNA sequence from +101 to +329 in the plasmid pKL201. These plasmids were then tested by assaying LUC activity after introduction of DNA into maize endosperm suspension cells by particle bombardment. Plasmid pBI221 containing the CaMV 35S promoter linked to a GUS gene was used as an internal control to correct for transfection efficiency. The results showed that levels of LUC expression driven by the two constructs were almost the same, suggesting that the first exon and intron region of Ae is not necessary for high level gene expression in maize endosperm cells.

To define the promoter sequences important for

- 53 -

Ae gene expression in maize endosperm cells via transient expression assays, a series of 5' deletion mutants were derived from the plasmid pKL201 using available restriction enzyme sites and PCR techniques (Fig. 5A,B).
5 The activity of each 5' deletion construct is presented in Fig. 5C. Two consecutive deletions of the Ae 5'-flanking sequence down to -1128 decreased the level of the LUC expression to approximately 40% of the full-length promoter level. However, the removal of an
10 additional 353 bp, to -775, restored LUC activity (Fig. 5C). This suggests that potent positive and negative distal *cis*-regulatory elements may be located in the regions from -2964 to -1129 and from -1128 to -776, respectively. Although further deletions down to -160
15 did not significantly change the levels of LUC expression, a dramatic reduction in the promoter strength was observed when an additional 111 bp, to -49, was deleted. This indicates that very strong positive regulatory element(s) are located at the 111-bp region
20 from -160 to -50. The presence of two GC boxes and one MRE motif in the region suggests the possibility that these conserved motifs may actually act as *cis*-regulatory elements essential for gene expression in maize endosperm cells.

25

EXAMPLE 3

Functional Analysis of *Sbe1* Promoter

30 We report in this example a functional analysis of the *Sbe1* promoter which reveals DNA sequence elements important for the high level and sugar responsive expression of the *Sbe1* gene in maize endosperm cells.

- 54 -

Materials and Methods:**Construction of Chimeric Plasmids. A**

transcriptional fusion of the *SbeI* promoter to a luciferase (LUC) reporter gene was made as follows. A BamHI restriction enzyme site was first created just before the translation initiation site of the *SbeI* gene by polymerase chain reaction (PCR): The DNA sequence between -253 and +27 of the *SbeI* gene was PCR-amplified with PI-1 and PI-2 primers (Table 3).

Table 3. Oligonucleotides used in PCR to create *SbeI* - LUC chimeric constructs.

	Primer ^a	Sequence ^b	Annealing Region ^c
15	PI-1 U	CCAGCTCCACGGTTGTTCTGT	-253 to -232
	PI-2 L	cgatggatccTGTGACGGCGTGTGAGTCCC	+8 to +27
	PI-3 L	agtcggatccTCAGGCGCACATTGCCGCCA	+209 to +228
	PI-4 U	gactgagctcATACCAAATGAAGCCAGGAG	+5382 to +5401
	PI-5 L	actggaattcGGAACAAGGAACGAAGAAAC	+5761 to +5780
20	PI-6 U	gatcaagcttACCAGCTCCACGGTTGTTCTG	-254 to -235
	PI-7 U	attcaagcttCAGATCCGGCTCAGGGTCAT	-196 to -177
	PI-8 L	TGCGACAAGGAGGGGGCCAT	-165 to -146

^a U and L indicate upper (sense) and lower (antisense) primers relative to the *SbeI*, respectively.

^b The lowercase letters designate restriction sites used for cloning.

^c Numbers represent distance relative to the transcription start site (+1) of the *SbeI*.

Four additional bases were included at 5'-end of the primers to provide for restriction enzyme sites at the ends of the PCR products for subsequent cloning. The bases were chosen randomly considering their effect on *T_m* value, dimer and stem-loop formation of the primers. -*Pfu* DNA polymerase (Stratagene, La Jolla, CA), which has proofreading activity, was used to enhance the fidelity of PCR amplification (*Pfu* DNA polymerase was used for all the following PCRs). Since an *ApaI* restriction enzyme

- 55 -

site (GGGCCC) is located immediately downstream of the 5' primer (PI-1) binding region of the *Sbel* promoter, -203 to -198, the PCR product was digested with *Apa*I and *Bam*HI. The resulting 236-bp fragment was then cloned into pBluescript SK⁻ and sequenced to verify that no misincorporation had occurred in the DNA sequence during the PCR amplification (all the following PCR products were sequenced). Next, the 236-bp fragment was ligated to the 1991-bp *Sal*I-*Apa*I *Sbel* promoter fragment and cloned into plasmid pLN cut with *Sal*I and *Bam*HI (promoterless LUC-NOS gene in pUC119) (Montgomery et al., Proc. Natl. Acad. Sci. USA, 90: 5939-5943, 1993), thereby creating plasmid pKL101.

To construct a translational fusion of the *Sbel* promoter containing the first exon and intron to a LUC reporter plasmid, the DNA sequence between -253 and +228 was amplified with the PI-1 primer and a 3' primer (PI-3) designed to anneal to the region just downstream of the first intron of the *Sbel* gene. The 493-bp PCR product was digested with *Apa*I and *Bam*HI, and the resulting 436-bp fragment was used to replace the *Apa*I-*Bam*HI fragment in pKL101. This construct was called pKLN101. To make pKLM101 which contains the *Sbel* promoter with four exons and introns, the 236-bp *Apa*I and *Bam*HI fragment in pKL101 was replaced with the 1816-bp *Sbel* genomic DNA fragment.

The plasmid pKLNS101 was derived from pKLN101 by replacing the nopaline synthase (NOS) 3' sequence with the native *Sbel* 3'-flanking sequence. To accomplish this, two primers PI-4 and PI-5 were designed to amplify *Sbel* DNA sequences containing the transcription stop signal and the polyadenylation site (from + 5382 to + 5780). A 419-bp PCR product was digested with *Sac*I and *Eco*RI, and the resulting fragment was then used for

- 56 -

substituting a 255-bp *SacI*-*EcoRI* NOS 3' sequence in the pKLN101.

To create a series of 5' deletions in the *SbeI* promoter, pKLN101 was first modified as follows: pKLN101
5 DNA was digested with *HindIII* and the resulting 7190-bp fragment lacking the 452-bp *HindIII* fragment was gel purified. The fragment was blunt ended by Klenow fill-in DNA synthesis and ligated with *SalI* linkers. After
complete digestion with *SalI*, the DNA fragment was
10 partially digested with *BamHI* to isolate the 1993-bp *SalI*-*BamHI* fragment, which was then gel purified and cloned into pLN cut with *SalI* and *BamHI* to produce pKLN101-1.

A series of 5' deletion mutants were made from
15 the plasmid pKLN101-1 using the S1 nuclease based Erase-a-Base system (Promega, Madison, WI) to produce the 5' deletion series plasmids, pKLN102 to pKLN107. All constructs were sequenced with pUC/M13 reverse primer to verify deletion end points. For the -254 and -196
20 deletion constructs, two regions of the *SbeI* promoter, -254 to -146 and -196 to -146, were PCR-amplified by primer PI-6 and PI-8, PI-7 and PI-8, respectively. Primers used in PCR to create the *SbeI*-LUC constructs are shown in Table 2. Since each 5' primer, PI-7 and PI-8,
25 contains a *HindIII* restriction enzyme site and a *BstXI* restriction enzyme site is located between -173 and -162, the PCR products were digested with *HindIII* and *BstXI* and the resulting fragments were used to replace the 2,047-bp *HindIII*-*BstXI* fragment of pKLN101.

30 **Linker-Scanning Mutagenesis.** A series of linker-scan mutations were introduced into the 60-bp DNA region from -314 to -255 as described by Kunkel et al (Kunkel et al., *Methods Enzymol.*, 154: 367-382, 1987).

- 57 -

Briefly, the HindIII-BamHI (-314 to +235) fragment from pKLN105, containing the DNA region to be altered, was subcloned into the corresponding sites of a M13mp19 vector to produce a single-stranded template. In order to increase mutant recovery efficiencies, the template was prepared from an *E. coli dut⁻ ung⁻* strain (CJ236) which allows the incorporation of uracil into the newly synthesized DNA. Next, a set of oligonucleotides with 10-bp mismatches shown in Table 4 were annealed to the template and extended with T7 DNA polymerase. After addition of T4 DNA ligase, the resulting heteroduplexes were introduced into a wild-type *E. coli* strain (MV1190) to generate mutated double-stranded DNAs. DNA sequencing was performed to verify that the desired mutations were correctly introduced and no unintended mutations had occurred.

To create the mutated *Sbe 1* promoter-LUC constructs (pLS1-1 to pLS1-6), the HindIII-BamHI fragment in pKLN105 was replaced by each mutated DNA sequence.

Table 4. Oligonucleotides used in linker-scanning mutagenesis.

Constructs	Oligonucleotides*
25 pLS1-1	CCCGGTTTGCCTTTTTTgCtgcaggacAAGCTTGGCGTAATCAT
pLS1-2	TTGCACGCTTCCCGGTTgaCtgcaggcTATTTTATGTAAGCTTG
pLS1-3	GCCTTTGGGCTTGCACGtcTagatagcTGCCTTTTTTTATTTTA
pLS1-4	GGGCCGATTGGCCTTTGactgcaggtactTCCCGGTTTGCCTTT
pLS1-5	AGCTGGTTCTGGGCCGAcgactgcagaGGCTTGCACGCTTCCCG
30 pLS1-6	ACAACCGTGGAGCTGGTgtcGactatcTTGGCCTTTGGGCTTGC

* The mutated bases are shown in lowercase letters, and restriction sites used for convenience of screening are underlined.

- 58 -

Particle Bombardment. Suspension culture cells of maize (*Zea mays*) endosperm (inbred A636) provided by J. L. Anthony (DEKALB Genetics Corporation, Mystic, CT) were grown in 250-mL large-mouth Erlenmeyer flasks
5 containing 80 mL of Murashige and Skoog basal salt medium supplemented with 0.4 mg/L thiamine 2 g/L asparagine and 30 g/L sucrose (Shannon and Liu, *Physiol. Plant*, 40: 285-291, 1977). The culture was maintained in the dark at 29°C on a rotary shaker (120 rpm) and was subcultured
10 every 7 days by transferring a portion of the cell suspension into fresh medium.

For particle bombardment, about 600 mg (fresh weight) of actively growing cells 3 days after subculture was evenly distributed over the surface of a piece of
15 filter paper (Whatmann #4, 55 mm in diameter) by vacuum filtration of 8 mL of suspension culture. The filter paper bearing the cells was then placed over three layers of filter paper (Whatmann #4, 70 mm in diameter) moistened with 5 mL of the liquid medium containing 12%
20 sucrose and positioned in the middle of a 10 cm petri dish.

60 mg of gold microcarriers (1.6 μ m particle size) were washed three times with 1 mL of 100% ethanol and twice with 1 mL of sterile deionized H₂O, resuspended
25 in 1 mL of sterile deionized H₂O, and dispensed in 50- μ L aliquots (3 mg/50 μ L). The *Sbe*I Promoter-LUC constructs and a GUS reference plasmid (pBI221, Jefferson, *Plant Mol. Biol. Rep.*, 5: 387-405, 1987) were co-precipitated onto the gold particles as follows: under continuous
30 vortexing, the following were added in order to each 50- μ L aliquot of gold particles: 5 μ L of DNA (8 μ g of LUC reporter plasmid, 4 μ g of GUS reference plasmid), 50 μ L of 2.5 M CaCl₂, and 20 μ L of 0.1 M spermidine (free base,

- 59 -

tissue culture grade). The gold particles coated with DNA were pelleted in an Eppendorf centrifuge at 10,000 rpm for 10 sec, rinsed with 250 μ L of 100% ethanol, and resuspended in 60 μ L of 100% ethanol. Immediately after
5 sonication, 8 μ L of the DNA-coated gold particles were pipetted onto the center of macrocarriers (Bio-Rad, Hercules, CA) and dried in a low humidity environment.

A Bio-Rad PDS-1000/He Biolistic Particle Delivery system was used for particle bombardment.

10 Bombardment parameters which were optimized include He pressure, gap distance (distance from power source to macroprojectile), and target distance (distance from microprojectile launch site to sample target). After optimization, all bombardments were performed in a dimly
15 light room at 650 psi under a vacuum of 26 inches of Hg with a distance of 10 cm between the cells and the barrel of the particle gun. Following bombardment, the petri dishes were sealed with Parafilm and then incubated in the dark at 25°C for 24 hr.

20 **GUS and LUC Assays.** The bombarded cells were harvested from the plates by vacuum filtration, frozen in liquid nitrogen, and ground with a pestle and mortar to a fine powder. The powder was then transferred into a microfuge tube and extracted with cell culture lysis
25 buffer containing 300 mM Tris-phosphate, pH 7.8, 2 mM DTT, 2 mM 1,2-diaminocyclohexane-N, N, N', N'-tetraacetic acid, 10% glycerol and 1% Triton X-100 (0.3 mL/g of tissue). Cell debris was pelleted in an Eppendorf centrifuge at 14,000 rpm for 10 min at 4°C and the
30 supernatant was split into two aliquots for assays of GUS and LUC activity.

For fluorometric GUS assays (Jefferson, Plant Mol. Biol. Rep., 5: 387-405, 1987), 30 μ L of the crude

- 60 -

extract was incubated at 37°C with 2 mM 4-methyl
umbelliferyl glucuronide in 0.3 mL of GUS assay buffer
(50 mM NaPO₄, pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 0.1%
Sarkosyl, 10 mM β-mercaptoethanol, 20% methanol). After
5 0, 1, and 2 hr of incubation, 0.1 mL aliquots were
removed and added to 0.9 mL of 0.2 M Na₂CO₃ to terminate
the reaction. A TKO 100 fluorometer (Hoeffer, San
Francisco, CA) calibrated by setting a 100-nM MU to 1,000
fluorescence units was used to measure fluorescence of
10 the product, 4-methyl umbelliferone (4-MU). For each
sample, results of GUS assay were plotted in a graph of
OD₄₀₅ (Y-axis) versus time in minutes and the GUS activity
was expressed simply as the slope of the line. GUS
activity from the maize endosperm suspension cells that
15 had been bombarded with the naked gold particles (no DNA)
was used as a control.

Using a luminometer (Monolight 1500; Analytical
Luminescence Laboratory, San Diego, CA), luciferase
activity was determined by measuring luminescence for 10
20 sec after mixing 20 μL of cell extract with 100 μL of
luciferase assay reagent containing 20 mM tricine, pH
7.8, 1.07 mM (MgCO₃)₄Mg(OH)₂·5H₂O, 2.67 mM MgSO₄, 0.1 mM
EDTA, 33.3 mM DTT, 270 μM coenzyme A, 470 μM luciferin
and 530 μM ATP. LUC activity from the maize endosperm
25 suspension cells that had been bombarded with the pLN
(promoterless LUC plasmid) was used as a control. To
correct differences in sample variability and
transfection efficiency, the luciferase activity (in
light unit) was normalized with GUS activity, yielding
30 the LUC/GUS ratio of each sample.

Nuclear Extract Preparation. Maize kernels
(inbred B73) were harvested 30 days after pollination and
frozen in liquid nitrogen. Nuclear extract was prepared

- 61 -

essentially according to the method described by Jensen et al. (Jensen et al., EMBO J., 7: 1265-1271, 1988).

Protein concentration was determined using a BCA protein assay kit (Pierce, Rockford, IL), according to the
5 manufacturer's instructions.

DNA Probe Preparation. The *SbeI* promoter region from -314 to -255 was PCR-amplified with a forward primer (5'-GGACTTACATAAAATAAAAAAGG CA) and a reverse primer (5'-TGCTAAAGCTTTCTGGGCCGATTGGCCTTTG) which contain
10 BamHI and HindIII restriction enzyme sites, respectively, at their 5' ends (underlined). The PCR product was digested with BamHI and HindIII, and the resulting fragment was cloned into pBlueScript SK⁻ cut with BamHI and HindIII to create plasmid pRb4-1. The plasmid
15 construct was verified with DNA sequencing. For electrophoretic mobility shift assays, the DNA fragment was cut out from the plasmid pRb4-1 with HindIII and BamHI, purified from agarose gels, and end-labeled with [α -³²P]-dCTP using the Klenow fragment.

Electrophoretic Mobility Shift Assay. The DNA-protein binding reaction was performed in 20 μ L of solution containing 0.5 ng of labeled probe, 10 μ g of nuclear protein, 1 μ g of poly (dI-dC).poly (dI-dC), 12% glycerol, 12 mM HEPES-NaOH (pH 7.9), 4 mM Tris-Cl (pH
25 7.9), 60 mM KCl, 1 mM EDTA and 1 mM DTT. After a 20-min incubation at room temperature, the samples were loaded into a 4% native polyacrylamide gel which was pre-run at 4°C for 1 hr at 150 V and electrophoresed for 2.5 hr at 150 V in Tris-glycine buffer at 4°C. Following
30 electrophoresis, the gel was dried with a gel dryer (Bio-Rad) and exposed to Kodak X-ray film with two intensifying screens for 24 hr.

- 62 -

Northern Blot Analysis. Total RNA was isolated according to Vries et al. (Vries et al., Plant Mol. Biol. Manual, B6: 1-13, 1988) from maize endosperm suspension cells which had been incubated for 24 hr in the MS basal salt media supplemented with 0.4 mg/L thiamine, 2 g/L asparagine and various amounts of sucrose from 0% to 15%. Northern blot analysis was performed as described in Gao et al. (1996, *supra*). Radioactivity was detected with a PhosphorImager and quantified with the ImageQuant software program (Molecular Dynamics). To correct for minor loading errors between the lanes, the blot was wash at 95°C in a 0.1% (w/v) SDS solution to remove the ³²P-labeled *Sbe1* cDNA probe and rehybridized with a ³²P-labeled tomato cDNA for 26S rRNA.

Results:

Transcribed Regions of the *Sbe1* Gene are involved in gene expression. To determine whether the 5' flanking sequence of the *Sbe1* gene has all of the DNA elements necessary to initiate transcription, a 2217-bp fragment upstream of the translation start site (-2191 to +27) was fused to the luciferase (LUC) reporter gene in pUC119 (pKL101) as shown in Figure 6A. The chimeric plasmid was then introduced into maize endosperm cells via particle bombardment along with a reference plasmid containing the cauliflower mosaic virus (CaMV) 35S promoter linked to a GUS gene (pBI221, Jefferson, 1987, *supra*) to correct for transfection efficiency. However, only very low levels of LUC activity were detected relative to other promoters previously described.

Since many reports indicated that DNA sequences within transcribed regions such as exons, introns and 3' flanking regions are involved in the expression of genes

- 63 -

in either qualitative or quantitative manners, three different types of translational fusion constructs were created to test the effect of downstream elements on *Sbe1* gene expression. First, the 5'-flanking sequence as well
5 as the first exon and intron of the *Sbe1* gene (-2190 to +228) were fused in-frame to the LUC reporter gene to make pKLN101 (Figure 6A, B). Second, the nopaline synthase (NOS) 3' sequence in the pKLN101 was replaced with the *Sbe1* 3' flanking sequence (399 bp in length),
10 which contains the translation stop codon and polyadenylation signal to create pKLS101. Finally, to determine whether an increase in the number of exon/introns enhances the gene expression, three more exons and introns from the *Sbe1* gene were added to the
15 pKLN101 to make pKLM101 (-2190 to +1617).

The results of transient expression assays using the chimeric constructs are shown in Figure 6C. Inclusion of the DNA sequence from +28 to +228 containing the first exon and intron increased the level of LUC
20 expression by 14-fold, suggesting that the first exon and intron region is required for high level expression of the *Sbe1* gene in maize endosperm cells. Since pKLN101 produce a fusion protein, however, we cannot completely rule out the possibility that the increase may be due to
25 changes in enzyme activity and/or turnover rate caused by the added amino acid sequences. If the additional amino acids have a negative effect, the enhancement of LUC activity observed would be greater than 14 fold.

Replacement of the NOS 3' end in pKLN101 with
30 the *Sbe1* 3' region did not have a significant effect on the level of LUC expression, implying that the *Sbe1* 3' untranslated region does not have indispensable control elements. However, it is still possible that the region

- 64 -

may be important for *Sbe1* gene expression in other cell types or inductive conditions.

Construct pKLM101 showed a slight reduction in LUC activity compared to pKLN101, indicating that
5 additional exons and introns had an adverse effect on LUC expression in maize endosperm suspension cells. The adverse effect could be explained by inefficient splicing, resulting from the introduction of multiple copies of the plasmid into a single cell, or by formation
10 of fusion protein consisting of the 5'-end of SBEI and luciferase, thus lowering LUC activity. Alternatively, it could also be actually due to the presence of negative *cis*-elements in this region.

5' deletion down to -314 did not significantly
15 affect the *Sbe1* promoter activity. To identify promoter sequences critical for *Sbe1* expression in maize endosperm cells, a series of 5' deletion mutants were derived from pKLN101 as shown in Figure 7A. The activity of each 5' deletion construct is presented in Figure 7B. Removal of
20 the sequences to -1332 caused a decrease in the level of the LUC expression, while deletion of an additional 422 bp, to -910, resulted in an increase in the activity of the construct. This suggests that potential positive and negative distal *cis*-regulatory elements may be located in
25 the regions from -2190 to -1332 and from -1332 to -910, respectively. Further deletions down to -315 did not significantly affect the promoter activity, but a severe reduction in the activity was observed when an additional 169 bp, to -145, was deleted. The -72 deletion construct
30 produced a level of the LUC activity slightly over background, showing that the minimal promoter is functional.

- 65 -

A 60-bp region is critical for the promoter activity. To further delimit sequences essential for high level expression of the promoter, two additional 5' deletions with about 60-bp intervals were created between -315 and -145. As shown in Figure 7, a deletion to -255 (pKLN108) severely reduced the expression of the LUC reporter gene, while a further deletion to -196 (pKLN109) did not further reduce promoter strength. This indicates that a strong positive regulatory element(s) is present in the 60-bp region between -315 and -255.

Linker-Scan Analysis reveals two *cis*-elements within the 60-bp region. Since the 5' deletion analyses indicated that the region of the *SbeI* promoter from -314 to -255 is critical for the promoter activity, the 60-bp DNA fragment was further dissected by oligonucleotide-directed *in vitro* mutagenesis, as described by Kunkel et al. (1987, *supra*). A series of six different substitution mutants, designated pLS1 to pLS6, were created by altering the wild-type DNA sequence of the *SbeI* promoter at 10-bp intervals. The mutations were made by creating transversion substitutions where possible, while at the same time introducing restriction enzyme sites for simplifying identification of the mutant forms.

The mutated constructs were tested for their promoter activity using the transient assay system, and the results of the experiments are shown in Figure 8. Mutations in the regions from -314 to -305 and -304 to -295, corresponding to pLS-1 and pLS-2, caused a decrease in the *SbeI* promoter activity to 60% and 72% of wild-type (pKLN105) expression, respectively. The pLS3 construct showed almost the same level of the LUC expression as the wild-type promoter, suggesting the nucleotides from -294 to -285 are not important for the promoter activity in

- 66 -

maize endosperm cells. However, mutation in the pLS-4 region (-284 to -275) decreased promoter activity to 40% of the wild-type level. Also, other two mutants, pLS-5 and pLS-6, resulted in a reduction of the promoter activity to 55% and 50% of the wild-type promoter, respectively.

The 60-bp fragment interacts with DNA-binding proteins. In order to investigate the possibility that a nuclear protein(s) might interact with the 60-bp *Sbe1* promoter fragment from -314 to -255, electrophoretic mobility shift assays were performed. The 60-bp fragment was ³²P end-labeled with Klenow fill-in reaction and then incubated with nuclear extract prepared from 30 days after pollination (DAP) maize kernels (B73), which have been demonstrated to highly express *Sbe1* gene (Gao et al., 1996, *supra*). Two major shifted bands were observed in the lane containing nuclear extract (lane 2) compared to the control (lane 1). The bands were not detected after inclusion of proteinase K in the binding reaction (lane 7), indicating the shifted bands represent DNA-protein complexes.

Competition assays were conducted to determine whether or not the complexes are due to the binding of sequence-specific proteins. Inclusion of 10-fold and 100-fold excess of the unlabeled 60-bp fragment in the binding reaction significantly reduced formation of the complexes (lane 3 and 4), while the same amount of non-specific competitor DNA failed to compete for binding (lane 5 and 6). Thus, the complexes appear to be the results of sequence-specific interactions between a nuclear protein(s) and the DNA fragment, consistent with the functional identification of this region as an important regulatory element. Using six 60-bp fragments

- 67 -

of linker scan mutants (LS-1 to 6) as competitors, we have found that LS-1 did not affect the intensity of the lower band, although the rest of linker scan mutants abolished its formation. This suggests that the lower
5 band may be the result of interactions between a trans-acting factor(s) and the sequence ACATAAAATA, located within LS-1. All of the linker scan mutants reduced the intensity of the slower migrating complex to varying degrees. Since LS-4, 5 and 6 were less effective
10 competitors, wild type sequences spanning these regions (-284 to -255) may be involved in formation of this complex, however, binding may involve several overlapping regions in this fragment.

Expression of the *Sbe1* gene is sugar-regulated.

15 The SBEs are expressed in a coordinate fashion with the granule-bound starch synthase (GBSS) and ADP-glucose pyrophosphorylase during maize endosperm development (Gao et al., 1996, *supra*). The ADP-glucose pyrophosphorylase gene (AGPase S) from potato and the genes encoding GBSS
20 and SBE in cassava plants have been shown to be induced by an exogenous supply of sugars (Muller-Rober et al., Mol. Gen. Genet., 224: 136-146, 1990; Giroux et al., Plant Physiol., 106: 713-722, 1994; Salehuzzaman et al., Plant Sci., 98: 53-62, 1994). This led us to speculate
25 that the *Sbe1* gene in maize may also be regulated by external sugar concentration.

To test this, maize endosperm suspension cells were incubated in MS media containing different concentrations of sucrose and their total endogenous RNAs
30 were analyzed by northern blot hybridization. Sucrose was used in preference to other metabolizable sugars, because it is known to be the major sugar unloading from the pedicel tissue of maize kernels. Increase in sucrose

- 68 -

concentration from 0% to 9% elevated the *Sbe1* mRNA level by two-fold, and at higher concentrations the increase was reduced. Hexoses such as glucose, fructose and myo-inositol also increased the level of the transcript in a similar fashion. However, L-glucose and PEG 200 at concentrations calculated to have the same osmotic potential as a 9% sucrose solution (263 mM) did not exhibit any effect, indicating that the response is not an osmotic effect but a sugar-specific phenomenon. These results suggest that expression of the *Sbe1* gene in maize endosperm cells is regulated by sugar availability like other starch biosynthetic genes (Giroux et al., 1994, *supra*). This metabolic feedback mechanism may serve as a system to fine tune the expression levels of *Sbe* genes relative to physiological status of a plant. In fact, Shannon et al. (Shannon et al., *Plant Physiol.*, 110: 835-843, 1996) showed that non-allelic starch mutants of maize accumulating high levels of sucrose in endosperm contain increased SBE activities compared to normal.

Since we determined that two *Sbe1* genes (*Sbela* and *Sbe1b*) with divergent 5'-flanking regions exist in the maize genome (Example 1), it was necessary to determine whether or not expression of the isolated *Sbe1* gene (*Sbela*) promoter is responding to external sucrose concentrations. To test this, a gene which is not regulated by sugar concentration was required for an internal control for the transient assay system. Since a CaMV 35S promoter has been used as a control in other studies investigating sucrose responsiveness of plant genes, the effect of sucrose on expression of the CaMV 35S promoter-GUS chimeric gene (pBI221) in maize endosperm cells was first investigated.

The plasmid pBI221 was bombarded into maize

- 69 -

endosperm suspension cells supplemented with 0% sucrose or 9% sucrose media and incubated at 25°C in the dark. After 48 hr incubation, GUS activity and protein concentration were measured from each sample to calculate specific GUS activity. The results showed that specific GUS activities of 9% sucrose samples were almost 2.5-fold higher than those of 0% sucrose samples, which is consistent with other reports. Since similar results were obtained from a ubiquitin promoter (pACH18) and -64 CaMV 35S minimal promoter which does not have an activation sequence (as)-1, a binding site for the transcription factor TGA-1a), it appeared that the elevated levels of expression by the CaMV 35S and ubiquitin promoters in 9% sucrose may be a general phenomenon simply due to an increase in energy source rather than a sugar-specific effect. Therefore, we reasoned that if the chimeric construct pKLN101 (the *Sbe1* promoter-LUC) is sugar-modulated, it will further enhance the level of LUC expression beyond the general increase at the higher sucrose concentration.

As shown in Figure 9, after normalization to GUS activity driven by the CaMV 35S promoter the plasmid pKLN101 still showed approximately two-fold greater LUC activity in 9% sucrose media than in 0% sucrose media. This is consistent with the result of the endogenous RNA analysis, indicating that the identified *Sbe1* gene is regulated by sugar availability. It also suggests that the nucleotide sequence containing a 2.2 kb 5'-flanking region and the first exon/intron of the *Sbe1* gene is sufficient to confer sugar responsiveness in maize endosperm cells.

Next, to delimit a region(s) necessary for the response, two deletion constructs, pKLN105 and pKLN106

- 70 -

were also tested in the transient expression system (Figure 9). Like pKLN101, pKLN105 (deletion end point -314) responded to a high sucrose concentration (9%) and increased LUC expression by approximately two-fold.

5 However, pKLN106 (deletion end point -145) showed similar levels of LUC expression in both low and high sucrose conditions. These results suggests that the region between -314 and -145 contains a *cis*-regulatory element(s) necessary for the sugar response in maize

10 endosperm cells. In addition, because the expression level was reduced in both + and - sucrose treated cells, other regulatory elements may also reside in this region.

Overexpression of mEmBP-1 protein represses the *Sbe1* gene expression. The canonical G-box sequence, CCACGTGG, was found in the 5'-flanking sequence of the

15 maize *Sbe1* (-228 to -221) as well as the rice *Sbe1* gene (-170 to -163). This evolutionary conservation suggests a possible role of the G-box motif in the regulation of gene expression, although our 5' deletion analysis did

20 not show it as an important regulatory element. It is known that the G-box motif resides in the promoters of many plant genes responding to a variety of different environmental and physiological stimuli and is often associated with additional regions which act as coupling

25 elements determining signal response specificity.

To test whether or not the G-box in the maize *Sbe1* promoter is interacting with a G-box binding protein in maize, mEmBP-1, which is a homologue of the wheat EmBP-1 (Guiltinan et al., Science, 250: 267-271, 1990)

30 and expressed during endosperm development, EMSA and DNase I footprint analyses were performed. As expected, the analyses clearly showed that EmBP-1 interacts with the G-box sequence *in vitro*. Since EmBP-1, a member of

- 71 -

the basic leucine zipper (bZIP) transcription factors, is implicated in ABA-induced *Em* gene expression in wheat.

The data prompted us to ask two questions. First, is the *Sbel* gene expression regulated by ABA concentration?

5 Second, can mEmBP-1 protein transactivate *Sbel* gene expression? Transient expression assays failed to show a relationship between exogenous ABA concentrations (1 to 100 μ M) and the *Sbel* promoter activity in the maize endosperm suspension cells, suggesting the G-box in the
10 *Sbel* promoter is not ABA-responsive.

To address the second question, a chimeric construct containing the cauliflower mosaic virus (CaMV) 35S promoter fused to the full-length mEmBP-1 cDNA (35S-mEmBP-1) was created and co-introduced with the plasmid
15 pKLN101 (a full-length *Sbel* promoter-LUC) into the maize endosperm suspension cells. We predicted overexpression of mEmBP-1 protein would enhance the LUC expression driven by the *Sbel* promoter, since mEmBP-1 is known as a bZIP transcription activator. Contrary to the
20 prediction, overexpression of mEmBP-1 protein actually resulted in a significant reduction (5-fold) of the *Sbel* promoter activity as shown in Figure 10. The effect was apparently selective for the *Sbel* promoter, since mEmBP-1 had little effect on expression of a LUC reporter gene
25 linked to the ubiquitin promoter (pACH18).

Interestingly, substitution of the G-box sequence (CCACGTGG) in pKLN105 with TTGA ACTA did not cause a reduction in promoter activity, suggesting that the G-box sequence is not required for high level expression of the
30 *Sbel* gene in maize endosperm cells.

Discussion

The expression pattern of the maize *Sbe1* gene has been investigated in almost all maize tissues (Gao et al., 1996). The *Sbe1* gene is constitutively expressed at a low level in vegetative tissues, while it is modulated during kernel development. Especially in the endosperm, *Sbe1* mRNA began to accumulate to high levels at the onset of rapid starch deposition. These findings suggest that the expression of *Sbe1* is regulated by certain factors which vary in concentration or activity during kernel development.

As a step toward understanding regulatory mechanisms controlling *Sbe1* gene expression, we analyzed the *Sbe1* promoter regions using a transient gene expression system. Transient expression assays showed that expression driven by the maize *Sbe1* promoter greatly depends on the presence of the DNA region spanning the first exon and intron of the maize *Sbe1*. Addition of the DNA sequence (+28 to +228) containing the first exon and intron of the *Sbe1* gene into the transcriptional chimeric construct (pKL101) increased reporter gene expression in maize endosperm suspension cells up to 14-fold. Since such DNA sequences containing transcriptional stimulating effects are useful in investigations of gene expression in plant cells and for plant genetic engineering, it will be necessary to determine whether or not the DNA sequence has the ability to increase gene expression under the control of other promoters. There are several examples of plant genes which are regulated by DNA sequences within the transcribed region. Among them, the first exon and intron sequences of the maize *Sh1* gene are one of the best examples studied so far (Vasil et al., Plant Physiol., 91: 1575-1579, 1985; Maas et al., Plant Mol.

Biol., 16: 199-207, 1991; Clancy et al., Plant Sci., 98: 151-161, 1994). The *Sh1* exon appears to have two separate *cis*-elements which act independently to increase gene expression via different mechanisms. One of the
5 elements may contain a novel promoter element which has the ability to interact with transcription factors binding upstream. The other acts possibly at the level of translation efficiency or mRNA stability. The enhancing effect of the *Sh1* intron is likely the result
10 of an increase in the level of mature cytoplasmatic mRNA level like the maize *Adh1* first intron.

5' deletion analysis of the maize *Sbe1* promoter revealed several *cis*-regulatory elements affecting the promoter activity in maize endosperm cells. Of special
15 interest was the identification of the 60-bp positive element located in the region from -314 to -255 relative to the transcription initiation site. Further investigation of the region using linker-scan analysis identified at least two separate regions, -314 to -295
20 and -284 to -255, which are critical for gene expression in maize endosperm cells.

Interestingly, the -314/-295 region has striking similarity with the sucrose-responsive element (SURE-1) of the potato patatin-1 promoter (Grierson et
25 al., Plant J., 5: 815-826, 1994), which has been shown to interact with a sucrose-inducible nuclear protein(s). Grierson et al. demonstrated that a 100-bp patatin-1 promoter fragment encompassing SURE-1 is sufficient to confer sucrose responsiveness. DNA sequences similar to
30 the -314/-295 region are also found in the promoter regions of other sugar-inducible genes, such as maize sucrose synthase (Shaw et al., Plant Physiol., 106: 1659-1665, 1994), *Arabidopsis* β -amylase (Mita et al., Plant

- 74 -

Physiol., 107: 895-904, 1995), and potato sporamin (Ohta et al., Mol. Gen. Genet., 225: 360-378, 1991). This finding along with the sugar enhanced expression of the *Sbe1* demonstrated by northern blot analysis and transient expression assay strongly suggest that the conserved sequences may be implicated in mediating sugar responsiveness of the *Sbe1* gene. This is further supported by our recent finding showing that the -314/-196 region of the *Sbe1* promoter is sufficient to confer the sucrose responsiveness to -64 CaMV 35S minimal promoter. Since high sucrose concentration media were used for the transient expression assays to maximize gene expression, it is understandable that mutation of this region decreased the level of LUC expression. It remains to be tested whether or not other *Sbe* genes are also sugar-modulated. To date, we have not detected sugar-dependent DNA binding activity associated with the *Sbe1* promoter.

In potato and cassava plants, sugars have been shown to regulate expression of genes involved in starch biosynthesis. Our results demonstrated that the maize *Sbe1* is also modulated by sugar concentration. Such sugar effect was not due to change in osmotic potential, because L-glucose and PEG which are osmotically active did not affect *Sbe1* gene expression. Recently, it has been reported that hexokinase is involved in sensing sugar concentration in higher plants, and sugar signaling mediated through hexokinase is uncoupled from sugar metabolism.

Sequence comparison between the rice (Kawasaki et al., 1993, *supra*) and maize *Sbe1* genomic DNAs (Example 1) revealed that the 5'-flanking sequences proximal to the protein-coding regions are highly divergent except

- 75 -

for the canonical G-box sequences (CCACTGG) which are located in similar positions relative to the corresponding transcription initiation sites. This evolutionary conservation between the species led us to postulate that the G-box may be involved in regulation of the *Sbel* gene expression possibly in response to one of the environmental or physiological stimuli, even though we failed to show the importance of the G-box in the *Sbel* promoter activity using the 5' deletion analysis. It is possible that G-box dependent mechanism controlling *Sbel* promoter activity could not be appraised in our endosperm suspension cultured cells. This hypothesis is supported by the results showing interaction of the G-box with mEmBP-1 protein *in vitro* and repression of the *Sbel* promoter activity by over-expression of mEmBP-1. Along with these, the finding that disruption of the G-box sequence (CCACGTGG) in pKLN105 did not cause a reduction in promoter activity led us to speculate that the G-box and its binding proteins are involved in down-regulation of the *Sbel* gene expression rather than up-regulation. Although a specific role for the G-box motif in *Sbel* gene expression has not been identified, there is a possibility that the G-box in the *Sbel* promoter may play a critical role under different environmental conditions or in different tissues.

It has been noted that mutations decreasing starch accumulation in maize endosperm also reduce storage protein synthesis, implying possible interactions between these pathways. It has been shown that mutations affecting synthetic events in one biosynthetic pathway affect expression of genes in both pathway, and demonstrated that expression of genes involved in starch and storage protein synthesis of the maize endosperm are

- 76 -

coordinately regulated. Elevation in sugar concentration or alteration of the osmotic potential of the endosperm was proposed to be a possible candidate for the primary signal triggering this coordinate expression. In this
5 context, knowledge of the *Sbe* promoter elements and their associated regulatory proteins may eventually lead to a better understanding of the regulatory mechanisms controlling all of the starch biosynthetic genes and the genes encoding storage proteins in maize endosperm.

10

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various
15 modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

-77-

We claim:

1. An isolated nucleic acid molecule for controlling expression of genes in transformed plant cells, which comprises a segment of an *Sbel* gene from a plant species selected from the group consisting of *Zea* spp., *Tripsacum* spp. and *Sorghum* spp., the segment commencing at a location about 3,000 bases upstream from a transcription initiation site of the gene, and ending at a location about 250 bases downstream from the transcription initiation site.

2. The nucleic acid molecule of claim 1, wherein the plant species is *Zea mays*.

3. The nucleic acid molecule of claim 1, isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:1.

4. A fragment of the nucleic acid molecule of claim 1, comprising a segment commencing at about 3,000 bases upstream from the transcription initiation site and terminating about 25 bases downstream from the transcription initiation site.

5. A fragment of the nucleic acid molecule of claim 1, comprising a segment located between about 25 and 250 bases downstream from the transcription initiation site, the fragment being capable of increasing promoter activity of homologous or heterologous promoters.

-78-

6. A fragment of the nucleic acid molecule of claim 1, comprising a segment located between about 145 - 315 bases upstream from the transcription initiation site, the fragment comprising elements that render
5 expression of the gene sugar-regulatable.

7. An isolated nucleic acid molecule for regulating expression of genes in transformed plant cells, which comprises a segment of a gene encoding a
10 *Sbe1* gene from a plant species selected from the group consisting of *Zea spp.*, *Tripsacum spp.* and *Sorghum spp.*, the segment comprising a 3' untranslated region commencing at a stop codon for gene's coding sequence, and ending at a location about 5900 bases downstream from
15 the gene's transcription initiation site.

8. The nucleic acid molecule of claim 6, wherein the plant species is *Zea mays*.

20 9. The nucleic acid molecule of claim 6, isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:1.

25 10. A DNA segment for effecting expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1, the segment comprising a promoter
30 and a transcription initiation site.

11. The DNA segment of claim 10, which further comprises an element included in a first exon of SEQ ID

-79-

NO:1, the element being capable of increasing promoter activity of homologous or heterologous promoters operably linked thereto.

5 12. The DNA segment of claim 10, which further comprises an element that confers sugar-regulatability on expression of the coding sequences.

10 13. The DNA segment of claim 10, isolated from a maize *Sbe1* gene.

15 14. A DNA segment for modulating expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1, the segment comprising a polyadenylation signal.

20 15. The DNA segment of claim 14, isolated from a maize *Sbe1* gene.

25 16. An isolated nucleic acid molecule for controlling expression of genes in transformed plant cells, which comprises a segment of a *Ae* gene from a plant, the segment commencing at a location about 3,000 bases upstream from a transcription initiation site of the gene, and ending at a location about 100 bases downstream from the transcription initiation site.

30 17. The nucleic acid molecule of claim 16, isolated from a *Ae* gene of a monocotylednous plant.

-80-

18. The nucleic acid molecule of claim 17,
isolated from a maize Ae gene.

19. The nucleic acid molecule of claim 16,
5 isolated from a gene having a coding sequence at least
60% homologous with the coding sequence defined by the
exons of SEQ ID NO:2.

20. A fragment of the nucleic acid molecule of
10 claim 16, comprising a segment located between about 50
and and 160 bases upstream from the transcription
initiation site.

21. An isolated nucleic acid molecule for
15 regulating expression of genes in transformed plant
cells, which comprises a segment of a gene encoding a Ae
gene from a plant, the segment comprising a 3'
untranslated region commencing at a stop codon for gene's
coding sequence, and ending at a location about 20,500
20 bases downstream from the gene's transcription initiation
site.

22. The nucleic acid molecule of claim 21,
isolated from a Ae gene of a monocotyledenous plant.
25

23. The nucleic acid molecule of claim 22,
isolated from a maize Ae gene.

24. The nucleic acid molecule of claim 21,
30 isolated from a gene having a coding sequence at least
60% homologous with the coding sequence defined by the
exons of SEQ ID NO:2.

-81-

25. A DNA segment for effecting expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by
5 exons of SEQ ID NO:2, the segment comprising a promoter and a transcription initiation site.

26. The DNA segment of claim 25, isolated from
10 a maize Ae gene.

27. A DNA segment for modulating expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by
15 exons of SEQ ID NO:2, the segment comprising a polyadenylation signal.

28. The DNA segment of claim 27, isolated from
20 a maize Ae gene.

29. A chimeric gene comprising a coding sequence operably linked to one or more DNA segments selected from the group consisting of:

a) an expression regulatory element,
25 isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1, the element comprising a promoter, a transcription initiation site and, optionally:

i) an element included in a first exon of
30 SEQ ID NO:1, capable of increasing promoter activity of homologous or heterologous promoters operably linked thereto; and

ii) an element that confers sugar-

-82-

regulatability on expression of the coding sequence;

b) a expression regulatory element,
isolated from a gene whose coding region hybridizes under
stringent conditions with a coding region defined by
5 exons of SEQ ID NO:2, the element comprising a promoter
and a transcription initiation site;

c) an expression modulatory element,
isolated from a gene whose coding region hybridizes under
stringent conditions with a coding region defined by
10 exons of SEQ ID NO:1, the element comprising a 3'
untranslated region; and

d) an expression modulatory element,
isolated from a gene whose coding region hybridizes under
stringent conditions with a coding region defined by
15 exons of SEQ ID NO:2, the element comprising a 3'
untranslated region.

30. The chimeric gene of claim 29, inserted
into a vector for transforming a cell.

20

31. A cell transformed with the vector of
claim 30.

32. The transformed cell of claim 31, which is
25 a plant cell.

33. The transformed cell of claim 32, which is
a maize plant cell.

30 34. A transgenic plant produced by
regenerating the transformed plant cell of claim 32.

-83-

35. A reproductive unit of the transgenic plant of claim 34.

5 37. A promoter isolated from a maize *Sbe1* gene.

38. A promoter isolated from a plant *Ae* gene.

10 39. The promoter of claim 38, isolated from a *Ae* gene of a monocotyledenous plant.

40. The promoter of claim 39, isolated from a maize *Ae* gene.

1/12

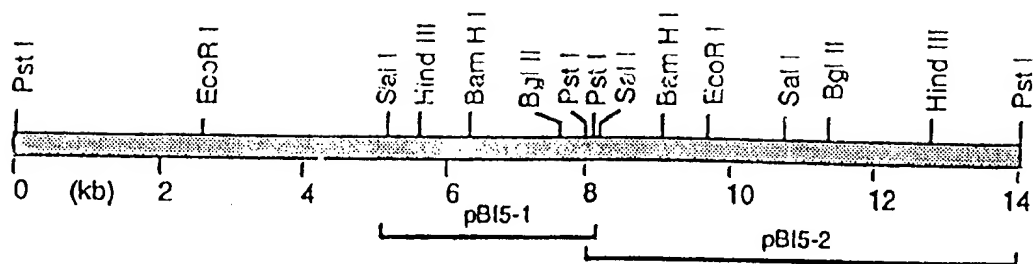


Fig. 1A

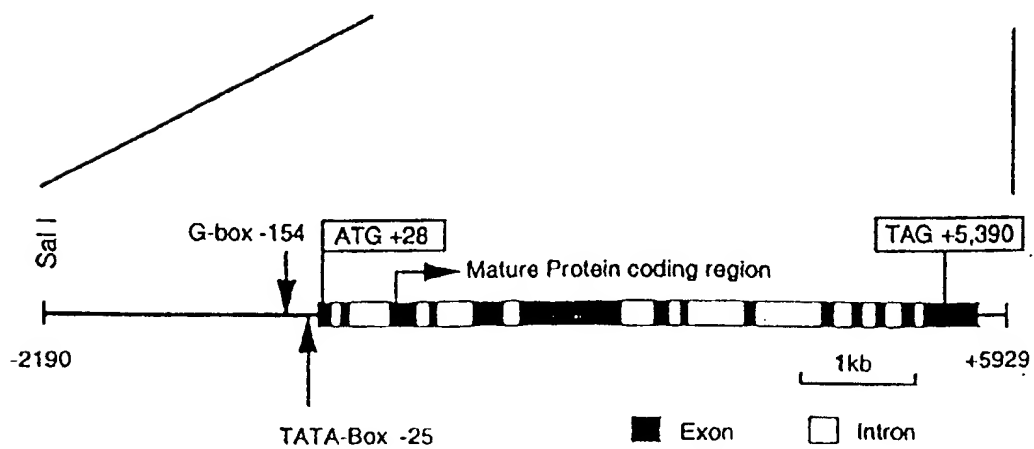


Fig. 1B

[illegible]

[illegible]

4/12

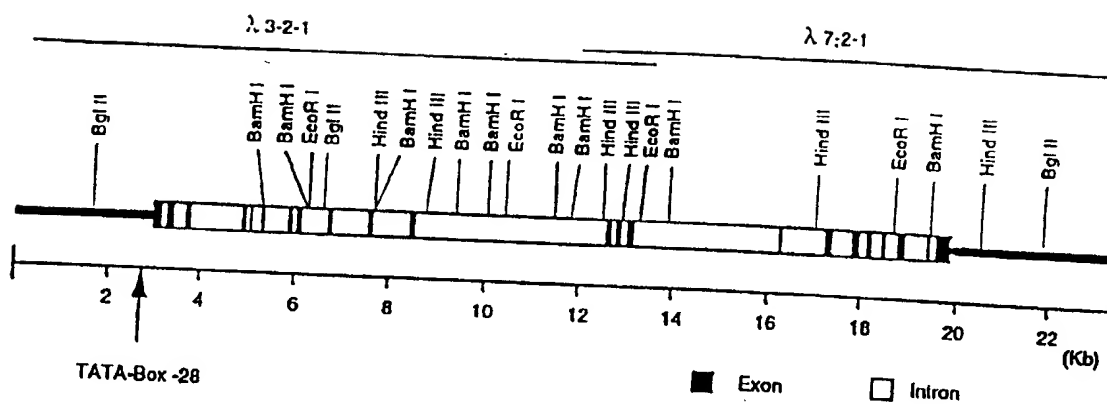


Fig. 3

[illegible]

[illegible]

7/12

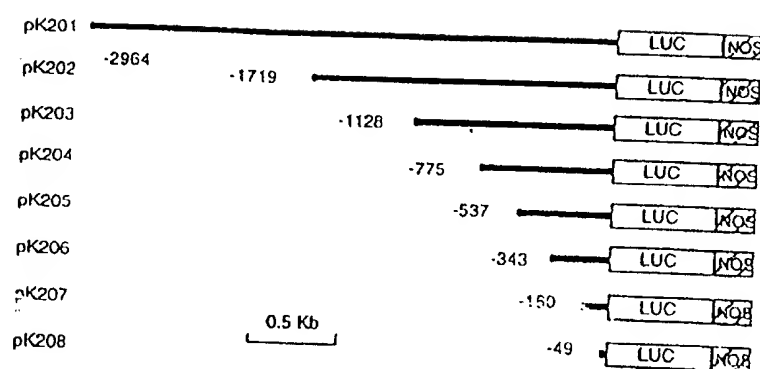


Fig. 5A

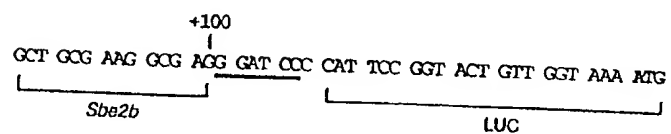


Fig. 5B

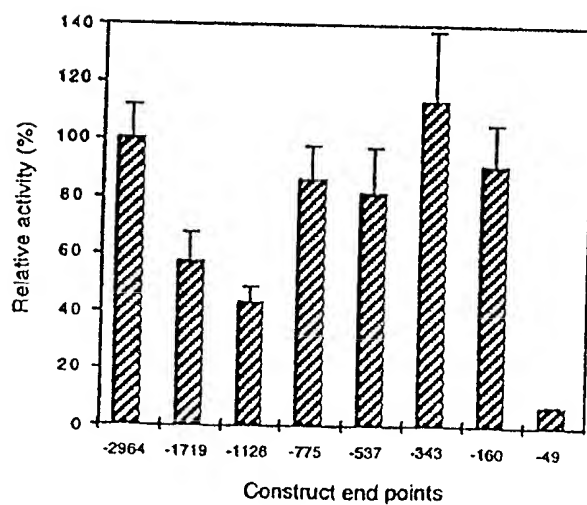


Fig. 5C

8/12

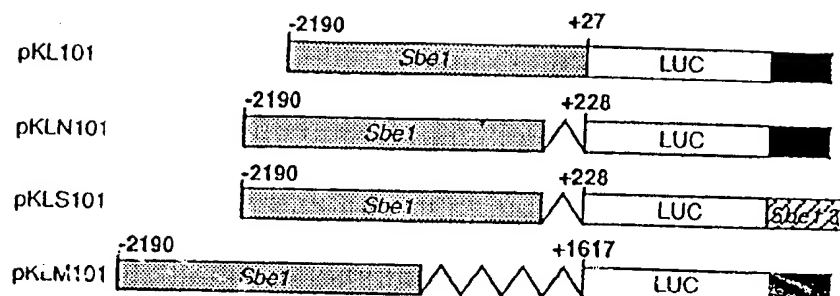


Fig. 6A

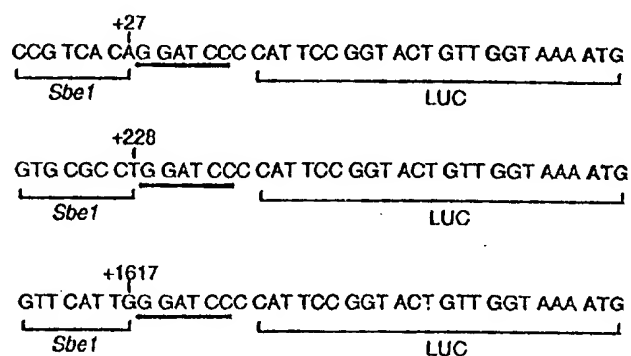


Fig. 6B

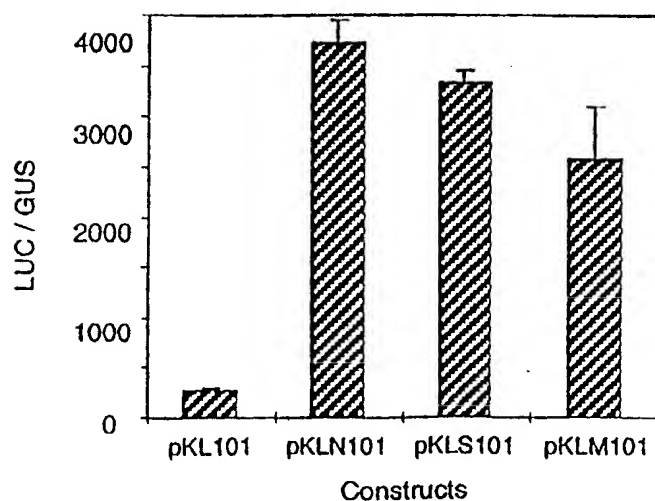


Fig. 6C

9/12

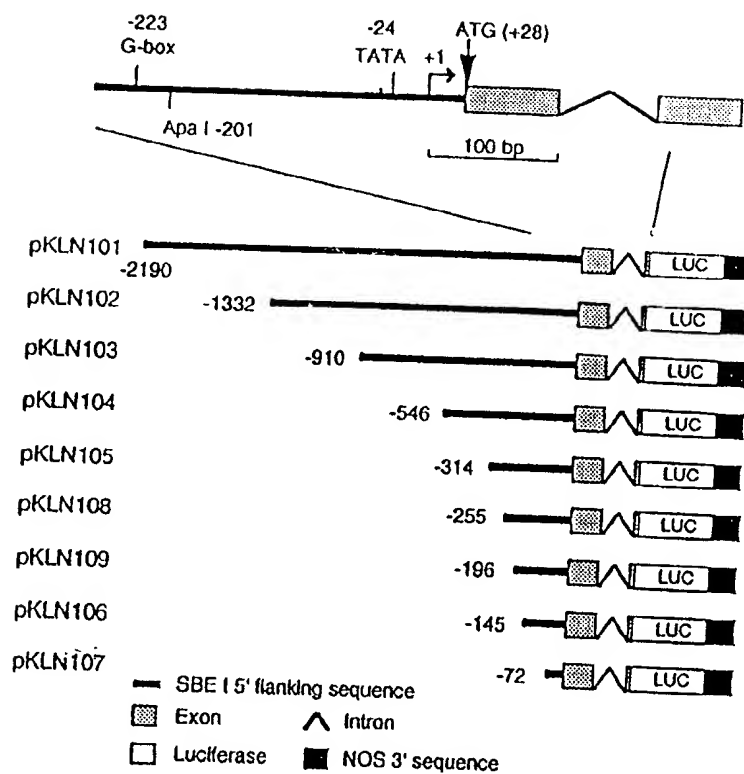


Fig. 7A

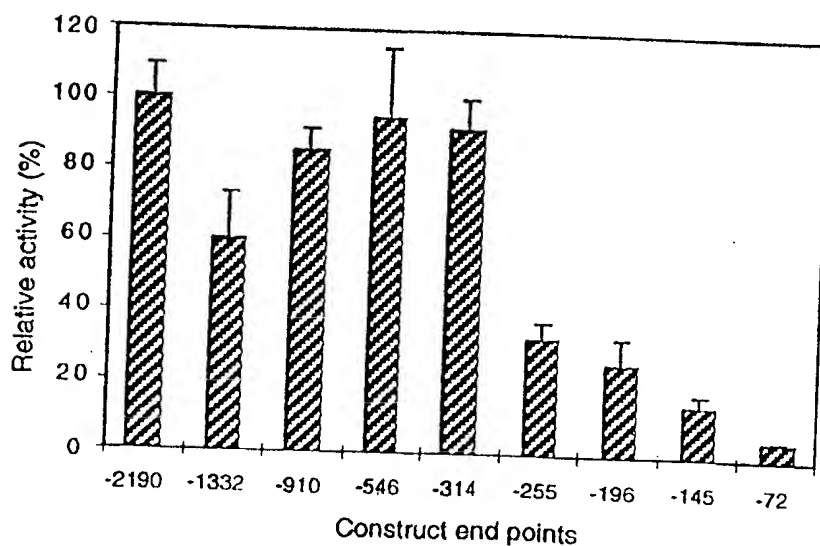


Fig. 7B

10/12

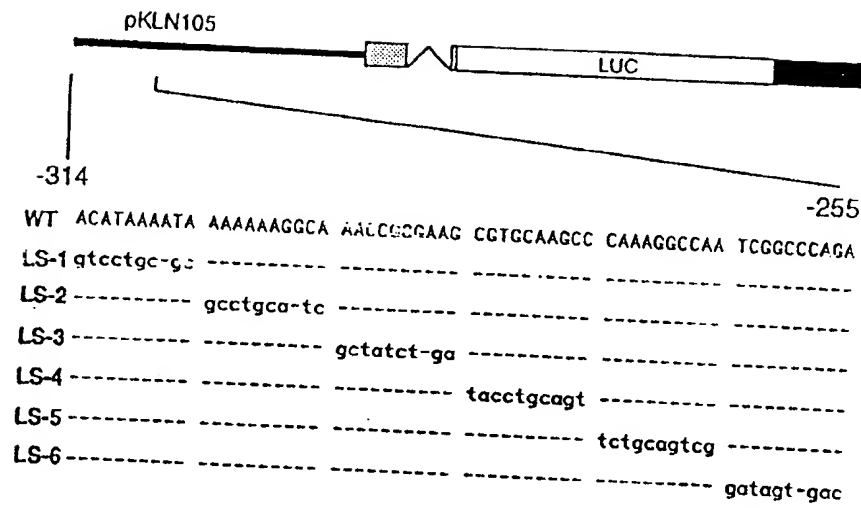


Fig. 8A

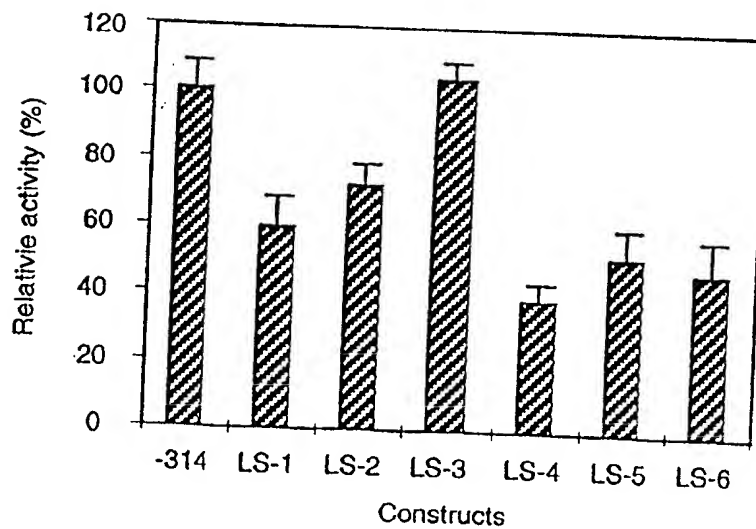


Fig. 8B

11/12

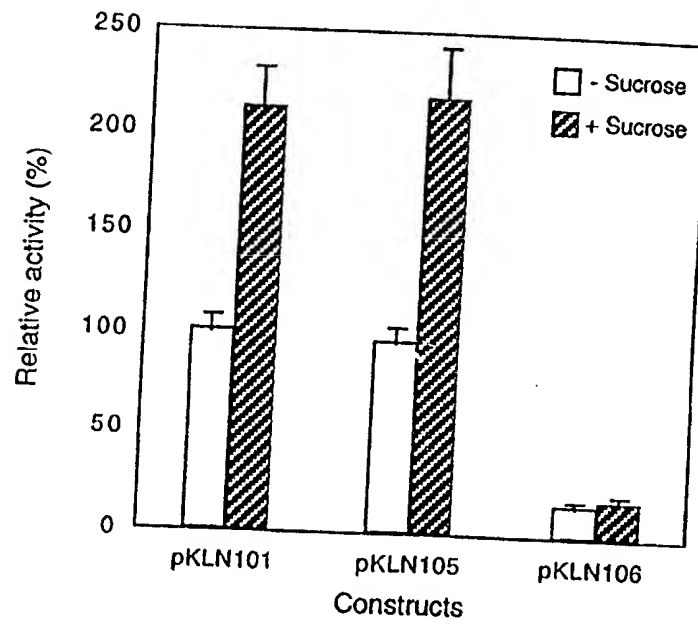


Fig. 9

12/12

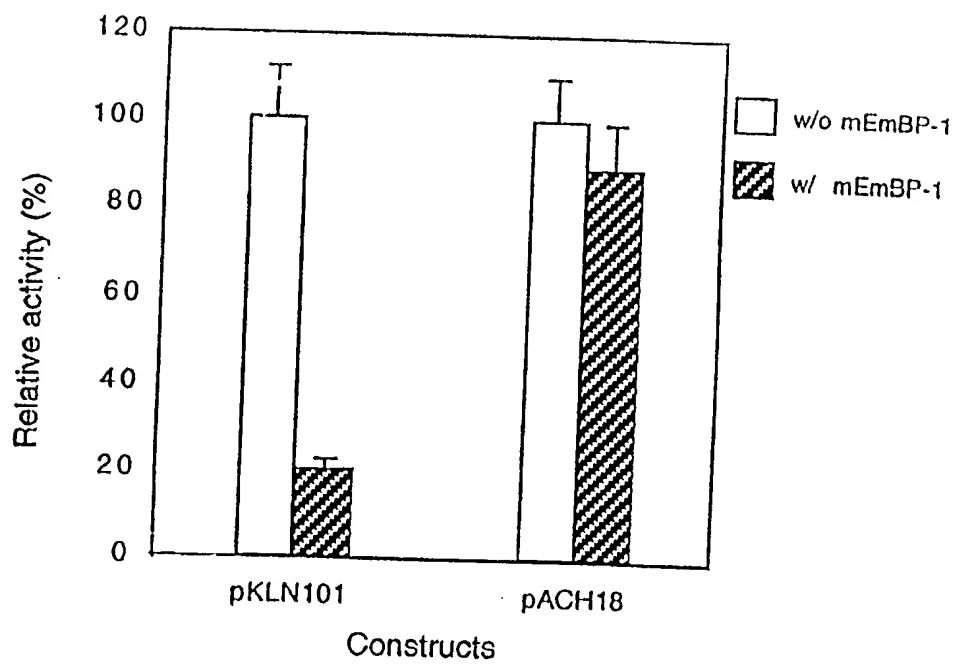


Fig. 10

1/13

SEQUENCE LISTING

<110> Mark A. Guiltinan
Kyung-Nam Kim

<120> Expression Control Elements from Genes
Encoding Starch Branching Enzymes

<130> PennState 1465

<150> US 60/089,049

<151> 1998-06-12

<150> US 60/089,050

<151> 1998-06-12

<160> 18

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 8119

<212> DNA

<213> Zea mays

<400> 1

gtcgactgcc	ctctagaccc	cgatgaagtc	aaggatagtg	agcgcccaac	caagtttagg	60
ctccccagc	cagatgatca	tgtaggtgt	tccactgcca	taaaagaaag	cgtccgtgat	120
gtcgtacttg	ggggcggtga	atacgtaaag	ggacagatgc	tcatatacct	tcatctgttt	180
tactcgaat	tactaatggt	ttccctatgt	cagctaaagt	gaaatctcac	ccgaccaatc	240
ttgcagattg	tgaatgactt	gaagtttacc	gtgaaccatg	ctgtggaacc	catcaatgaa	300
aagctgcaca	tgatatccga	gaacatcaaa	aagcgtgaga	agggaaagag	gaaaacgaat	360
gatgatagct	gcatcagttc	accaacatcc	ttgaccaggg	tcatcagtg	gattgatgat	420
ccgctcctga	acggaagctt	gagtgacaac	agcgggtcaa	agaaagcacg	gaggcagcgc	480
aggaaatcag	gctatgcaag	cgagagagc	ggtggggaga	gcagcgacca	tggcctaggc	540
gggtttgaga	tccgagggaa	cagatggatt	actagggagt	gacagctagg	agcaatctgt	600
ggagatacta	tgtaaatgtc	acaggagttg	gcctgtggtg	atgtgcgata	accgggccag	660
tgggttctga	gcgtaagacg	aggtctctag	caatgatatg	ttgtagcgtg	actaaatcga	720
gatgagcaat	caggttgttg	tctgttgatt	ctttcggctg	accatttcgg	ctgctgaggc	780
cttgatcctg	tatttgagat	gtcaggccta	cttagagcaa	atgctggaag	gaagcaatgt	840
gcctgttctg	aggtccctcc	agcatcactt	gtaatttgta	aagtactcaa	gtacagacta	900
gaactagtca	tgcataccag	aagttgaggc	ctgagatgcc	gtgccactct	atttgctctg	960
tttcatgctt	tgccgttatt	atatcactat	tttactgtt	cattctaagg	ccgttggtcac	1020
acgggtacca	gaagagcaga	acaccttttt	acgttccttt	tgctgctatg	gatggacttc	1080
gagagtccct	tgagaaagca	ctctcatcta	tctttccaat	tccttgtttc	cttggtggatc	1140
ctgatcactg	aagatggcca	tttgattttt	cacttttcag	ttatatataa	agtgaagacg	1200
gttgctctac	gcaatatgtg	aagtgaaga	tgagctctac	actgcaagga	acccaagaa	1260
aaaaccacag	gtaagtttgt	ctgattgaca	acggctagag	ctgtctgacg	atgcagatga	1320
ctccgcatgt	tcattcatca	ctttaatttg	ttagaaaagc	tacacaaaaa	gcacatcaga	1380
aaaaaagaag	aaaaagctat	ttagtggtac	tactagttgc	aacttgcaat	aatgatgatg	1440
ataaatctgc	acaagccata	gctatgctat	atgctatagc	tatgtatatg	tacacaaaaa	1500
atacattttt	tgtgctat	tttttaccgc	tagtataata	tccatgtctt	gctacaacac	1560
acaatcatat	ttaataccta	taaaaataaa	atttaatat	aaaataaaca	tatggtccac	1620
accatatatt	aaactgctaa	aaacaaatat	tataactcat	ttgatcgttc	atcctctttc	1680
ggttagttag	ttggacagtg	agagcgctgc	atcgtgttat	tgggttgac	tggtttctca	1740
cggctcatct	gtgttgtaac	gacctatcta	tggtcataaa	aactattagg	attattgtta	1800

ggcgaaaaat	gagggagata	aataaaccta	taagcaggac	acatgaaaca	tatgcttttaa	1860
acagtagaga	ttaacaacat	aaaataaaaa	aaggcaaac	gggaagcgtg	caagcccaaa	1920
ggccaatcgg	cccagaacca	gctccacggt	tgttcgtgtc	cgcccacgtg	gcacgcccgg	1980
ccattccggg	cccacagatc	cggctcaggg	tcattgtgcca	ctgccatggc	ccccctccttg	2040
tcgcagcggc	agattgcgac	ggggaagaaa	ggtgaggaga	ccaagcgaaa	aaaatcacgc	2100
tttcattgcg	aggggggagg	atagggagga	agacgccaa	ccagctccag	tccggcaccc	2160
gatataaagc	ggcaggcact	tggattgctg	acgagatggg	actcacacgc	cgtcacaatg	2220
ctgtgcctcg	tgctgccttc	ttcctcgccg	actccgcttc	cgcgcgcgcg	gcgctctcgc	2280
tcgcatgctg	atcggggcggc	accgcggggg	atcgcggtaa	gctgcggcga	tccgggggct	2340
ggtggattta	ctacttttgc	ttttttttct	tctgtactga	tggttcatag	tttcgcaggg	2400
tggcggcaat	gtgcgcctga	gtgtgttgte	tgtccagtg	aaggctcgcc	ggtcaggggt	2460
gcggaaggta	gatctttccc	cagattagaa	cagtttgta	cggacaaaat	ggaatttggt	2520
agtagttttg	gagagagaaa	aaaaactctg	gaatttggtg	attgggtcct	gctcgctttg	2580
gatttggtct	tagtgctttg	ttatcagtag	taatcatttt	accaagtcta	gcgcatctca	2640
gatttcaact	gaagttgttt	aaattttggt	ggttggtttt	cttcagccat	cctaattcatt	2700
tattacataa	taattagttg	aaataaagtg	ctatatattga	tattttctgtg	gttgctgcag	2760
cagctaatta	gcttctactac	attttcacgt	gttagcttct	attgctcgtg	taattctgaa	2820
tctggcttta	acgcactgtg	caaggtcaag	agcaaatcgc	ccactgcagc	tactgtgcaa	2880
gaagataaaa	ctatggcaac	tgccaaaggc	gatgtcgacc	atctcccat	atacgacctg	2940
gaccccaagc	tggagatatt	caaggaccat	ttcaggtacc	ggatgaaaag	attcctagag	3000
cagaaaggat	caattgaaga	aaatgagggg	agtcttgaat	ctttttctaa	aggttaggct	3060
taatacattc	aaatgcta	acaaggacca	tttcacaata	tttatagcac	ctgcccattgc	3120
cgcaaaatat	tatctgcac	atcaacattt	ccggattttg	tatttttttc	aggctatttg	3180
aaatttgga	ttaatacaaa	tgaggatgga	actgtatatc	gtgaatgggc	acctgctgcg	3240
cagtaagtgc	taatgttgtc	atgcaaacat	gatgtactgg	cgggggtatcg	ttttttccca	3300
ttttgcttgt	gaagagatat	cgtctgatag	taactatatt	aaaaaaaaag	aacagctaaa	3360
acacaggaga	gtaatcaatg	tttgggagta	gaaaactgtt	cggatatttc	attcttgtgc	3420
aagactgtca	tgtcttagct	gaggaggcta	ttttttttct	tgctgtgatt	acataatgct	3480
gttttcttct	tatgacaata	tactttccac	tattttaata	gattgatgca	tttgacttga	3540
gttttttact	catgggtgta	gggaggcaga	gcttattgtt	gacttcaatg	actggaatgg	3600
tgcaaaccat	aagatggaga	aggataaatt	tggtgtttgg	tcgatcaaaa	tgaccatgt	3660
caaagggaaa	cctgccatcc	ctcacaattc	caagggtaaa	tttcgctttc	tacatggtgg	3720
agtatgggtt	gatcgtattc	cagcattgat	tcgttatcg	actgttgatg	cctctaaatt	3780
tggagctccc	tatgatgggt	ttcattggga	tcctcctgct	tctgaaaggt	ctctttctac	3840
tgtcctggac	cacattgtaa	acttaagcaa	tagaccttag	cagacaaaaa	tatatgacaa	3900
aattatagta	ccttttacct	ctggtttgca	ggtacacatt	taagcatcct	cggccttcaa	3960
agcctgctgc	tcacgctatc	tatgaagccc	atgtaggat	gagtgggtgaa	aagccagcag	4020
taagcacata	tagggaattt	gcagacaatg	tggtgccacg	catacgagca	aataactaca	4080
acacagttca	gttgatggca	gttatggagc	attcgtacta	tgcttctttc	gggtaccatg	4140
tgacaaattt	ctttgcgggt	agcagcagat	caggcacacc	agaggacctc	aaatatcttg	4200
ttgataaggc	acacagtttg	ggtttgcgag	ttctgatgga	tggtgtccat	agccatgcaa	4260
gtaataatgt	cacagatggt	ttaaatggct	atgatgttgg	acaaagcacc	caagagtcct	4320
attttcatgc	gggagataga	ggttatcata	aactttggga	tagtcggctg	ttcaactatg	4380
ctaactggga	ggtattaagg	tttcttcttt	ctaacctgag	atattggttg	gatgaattca	4440
tgtttgatgg	cttccgattt	gatggagtta	catcaatgct	gtatcatcac	catggtatca	4500
atgtgggggt	tactggaaac	taccaggaat	atttcagttt	ggacacagct	gtggatgcag	4560
ttgtttacat	gatgcttgca	aaccatttaa	tgacacaact	cttgccagaa	gcaactgttg	4620
ttgctgaaga	tgtttcaggc	atgccgggtc	tttgccggcc	agttgatgaa	ggtgggggtg	4680
ggtttgacta	tcgcctggca	atggctatcc	ctgatagatg	gattgactac	ctgaagaata	4740
aagatgactc	tgagtgggtc	atgggtgaaa	tagcgcatac	tttgactaac	aggagatata	4800
ctgaaaaatg	catcgcatat	gctgagagcc	atgatcaggt	acctgcatt	atataataac	4860
tttataagga	agtatagttg	gatgcagcaa	agtattttat	ttccttgttg	ggtagacaac	4920
acatggggca	agtatctttt	aatcagacag	tactatat	catcacattt	ctatcctgat	4980
tgctagtaac	tggattagtt	attcataaaa	aattaggccc	tcctaagcat	gtttaagaat	5040
ttgcaatcac	atatgacctg	tttatcttat	actggaaatt	aaaatcatgc	aaattttcag	5100
tctattgttg	gcgacaaaac	tattgcattt	ctcctgatgg	acaaggaaat	gtacactggc	5160
atgtcagact	tgacgcctgc	ttcacctaca	attgatcgag	ggattgcact	ccaaaagggt	5220
ccattctcct	caagttctgg	ttgaacctgg	ttttcaataa	ttcttgaatg	agttccctca	5280
atataccttg	cttattgtgt	tctgtacact	ccagatgatt	cacttcatca	caatggccct	5340

tggaggtgat	ggctacttga	attttatggg	aaatgaggtg	aaatctcggg	cttttgaaaa	5400
atgtttccca	ttaataagcc	tgcataaacc	ttttgtattt	tttttctatt	tgaggtttca	5460
gcatttagag	tcgaccctct	attttctcag	taactgaaat	tcctaattgg	tatttgaatg	5520
gttaatgtta	attccgagta	gcgctcgttg	cacatgaaaa	tgtggtctgt	ttatggctgc	5580
taatcctttt	gtgacttgtc	caatatactc	gtacaacatg	ggatacatgt	aacccttcac	5640
tatcaagcat	cgcttaaatg	tcaacatcat	acagagatat	tatcaagtct	ttctcctgca	5700
ctatcctgga	atgaaaaaaa	agagttgttt	catatcgtat	ctgtactttt	acaaatgagt	5760
gtgatctttg	gagaagttag	ttgacggaat	gggtccactgt	attttgtttg	ttatatgttt	5820
ctccttattc	atagtgtatt	caattgatgt	gcaattctct	tctagtttgg	tcaccagaa	5880
tggattgact	ttccaagaga	agggaacaac	tggagctatg	ataaatgcag	acgacagtgg	5940
agccttggtg	acactgatca	cttgccgtac	aaggttatgt	ctatgaatgc	aatccttata	6000
agatttttgt	tctggcacca	ttctaggctc	tctttgacct	ttacctcttc	catttatcgc	6060
tgtctcctat	actagctaga	tctgtctctg	agactagcta	gatgtcagtt	ccttactact	6120
attacactct	tgagttgagc	tgggtccaga	tatgcttttc	cacactttgt	ttgtctgccc	6180
cactgcccta	ctatagtaaa	ttgcaccaac	cccgtagatg	ttactcaata	aacagggttg	6240
attggcctgt	gcaaacctag	tattttgtaa	gttgacctgg	ctttatgtat	tgccattgtt	6300
gaataaaaaa	aattttgatg	cattgattgg	tatgtcattt	tcataacttt	gcagacctag	6360
gaatgtactt	ctcaatcatt	ggttgttttg	ttgatagaaa	ttattctgtt	tacagcatat	6420
ggtggactaa	atatccaagt	gaacgtttgt	ttgtctgttt	tttttttcta	ccctcaagt	6480
ttacaattta	tcccggctaa	catattaatg	ttttctcctt	ggctctaata	ctgaacagta	6540
catgaatgcg	tttgaccaag	cgatgaatgc	gtcgcgatgag	agattttcct	tcctttctgc	6600
gtcaaagcag	atcgtcagcg	acatgaacga	tgaggaaaag	gtaaggattt	cagaataact	6660
ggtatgagag	tggttgcgacg	tttctttctt	tcttaggaag	agtggtaact	ctagctttgt	6720
tgctgtattt	tgccgattaa	ggcgagcttc	aatttaggag	tggatgggtt	gtcagcatta	6780
ttttttctac	ttcgtttcag	gttattgtct	ttgaacgtgg	agatttagtt	ttgtttttca	6840
atttccatcc	caagaaaact	tacgaggggt	aagtcaccag	ttgtaaaaacc	ctgtcttttc	6900
agagtccctt	ttgcctgtgg	tataatataa	tatagtgtgc	ttacctccca	tctctgtctc	6960
gcaacttata	tcagactaca	aagtgggatg	cgatttgccct	gggaaataca	gagtagccct	7020
ggactctgat	gctctggtct	tcggtggaca	tgaagagta	agtagtgacg	gtagacgctg	7080
aaagatgttt	ttttttttta	atttttgccc	tactcccgtg	gttggggccg	gccggccgct	7140
gagaatcgat	catatcaggc	tgtcgtgttt	gcgagactca	tggatgctgt	gacgcagggt	7200
ggccacgacg	tggatcactt	cacgtcgcct	gaaggggtgc	caggggtgcc	cgaaacgaac	7260
ttcaacaacc	ggccgaactc	gttcaaagtc	ctttctccgc	cccgcacctg	tgtggtaatg	7320
ttcacttact	ctcagctgaa	ctgatgagct	agaatgtatc	cgccctgaca	aaccgtcctg	7380
ctgccaggct	tattaccgtg	tagacgaagc	aggggctgga	cgacgtcttc	acgcgaaagc	7440
agagacagga	aagacgtctc	cagcagagag	catcgacgtc	aaagcttcca	gagctagtag	7500
caaagaagac	aaggaggcaa	cggctggtgg	caagaaggga	tggaaagttg	cgcggcagcc	7560
atccgatcaa	gataccaaat	gaagccagga	gtccttggtg	aggactggac	tggctgccgg	7620
cgccctgtta	gtagtctctg	tctactggac	tagccgcccg	tggcgccctt	ggaacgggtc	7680
tttctgttag	cttgcaaggc	actggtgtct	catcacccag	caggcaggca	ctgcttgtat	7740
agctttttcta	gaataataat	cagggatgga	tggatgggtg	gtattggcta	tctggctaga	7800
cgtgcatgtg	cccagtttgt	atgtacagga	gcagttcccc	tccagaataa	aaaaaaactt	7860
gttggggggg	ttttcttcta	ctctgtagtc	tgttcttgct	aggttgacga	agatgtttga	7920
tattagatgc	catagatcat	gtactgttaa	gtttcttcgt	tccttgttcc	ctgtccagtt	7980
cacatttgat	tccagctggt	cagcaggccg	gtcagctcag	ctccacaccc	ggggccaggc	8040
cggcntcacg	caggatcttt	tcatttcatg	cgtcataaca	caaacacttt	tgatcttttc	8100
agacaaaaaa	gatggatcg					8119

<210> 2

<211> 23449

<212> DNA

<213> Zea mays

<400> 2

ctcgagaatt	atttttgata	atccatcgga	ccgtcattgt	tcattggtcag	tgccacacaa	60
gtccggaccg	cctccaaagg	atagtctttg	gcgtgcgcgg	gtgagccaac	tggttagctga	120
cgtggcacc	aatccagagg	tgactagac	taatccggtg	cccgtcagcg	aaggaaagtg	180
tcaaccagaa	tcctggactc	atgacgcag	aaccgatcca	gtacccccgc	aagtaagctg	240
atatttgaca	tctttttgca	agaaaggagc	aatgactatg	gggtcttttg	gagctataaa	300

agcaccceca	tggcgacctc	cttcagaacc	caagcactcc	aagagtgtca	catcactccg	360
actctcttct	gtaactcatt	ctagtgattt	agtgagatca	atgcgctcatt	tctgagctgt	420
tcttgtgtat	gtgtgacttt	gcgcttttgt	gatcttctct	cttgagtttt	gctgctacgt	480
tttggctttg	tatttgtaat	ctctcccttc	tcttgctttg	agttgtaact	ctgatattgt	540
gtacgactgc	aagagactct	aaattatgga	gattccttgc	aaacagggat	ataagtgata	600
aggaaggacg	tggcactcaa	tttcgatctt	tggatcactt	gagagcggct	gattgcaatc	660
cttgtcgttg	gcaaccacaa	cgtgaagtag	tcaagcattt	taggcttgac	cgaaccacga	720
gaaaaatcat	cgtgtcttgt	gtgcttatcc	acttgtgatt	tttctccctc	tttctaagtc	780
tctcagattc	acttgaataa	ttactcttaa	tatattatat	atcttgaaag	agcaaacaaag	840
ttgaagatga	cttactttct	tcttctctct	atttaaactt	ggcttggttt	cactaacttc	900
catttccaga	ccaaacttgt	gttttttagt	ttgtttttgc	agggtcacct	attcaccctc	960
tctaggtgct	ctcaagaggt	gttagctatg	tttccctagg	cgtgtggcgg	tcaccttttc	1020
gatagatatg	aataagttcg	aactttcttt	agaaacaata	agtcgcaa	tcaacaaagt	1080
actaaatgcg	ctatacaaca	tgtctagtga	cgtaatgaag	ctaaaggacg	accaatttctg	1140
tgagattcac	cctagactat	tggaggtgcg	tttttggcaa	cattgtaggg	attgtataga	1200
actatcgatg	gaagccactt	tccgtgaatc	gtacctaggc	ggcatgtata	cgcgtcgcaa	1260
aatgtcatgg	cagtatgcga	cttcgatatg	tggttttacct	ttgccgtgac	tggttgggtg	1320
ggctcagttc	atgatctgta	ggtactacaa	gagactctca	ttgccacgac	aacatctttc	1380
cccatccacc	tgaaggtaat	gttttgggcg	aacaaaaata	attgtcacaa	atgttataaa	1440
tttgaatagc	attgtctaat	aacttgtttt	ggtaatgcat	gaatatatta	cttggctcgat	1500
tccggtcagc	gtaacaggtt	gggctacttt	gccccgtata	agggacaaaa	gtatcacgtt	1560
acagaacgac	aacatgggag	acatcccgtg	gttgagaaag	gggtattcaa	ctatgcacat	1620
tcacccctaa	gaaatgtcat	agagcgttca	tattgtgtgt	ttgaaggatga	agtggcagat	1680
cttgcttagt	ctcccatgtt	tctcgggtcg	gaagcaataa	aaaataatta	taacttgtat	1740
gaccctacat	aactttatta	gggtagtgct	tatacacaa	cacgacttgg	aaattttgtg	1800
tctgggaagc	atgtaactga	tgtagggtga	acagaaacta	gtagtggacc	attaggtgag	1860
ttagacacga	gtgagtttcc	agatgacatt	tatgctgcgc	tagtgacatg	acctaccatt	1920
tgtatttcat	agatgtggca	tgaacatgtt	tatgttgtaa	tgttctacgt	cagtgcatta	1980
tgtaatgtgt	taattaattt	aaggttgccc	tattagcaat	gaaatacaaa	tctccatcgg	2040
gatcggaggg	aattaaggtg	aaaatgaact	tatttctctc	ccaacccctt	caatctcgaa	2100
ggggatttga	gtttccaaac	tagatcctaa	aagcaagtat	aataaggtga	tgtaggcgag	2160
ctgtaggata	taacacatca	gatttgtgat	gatatgaaag	aaaaaaaaatg	aagagagaat	2220
gaaagcgaac	tgttgctcac	ggactctaag	atataatata	agagggaagag	atgagctaag	2280
tattaaatgt	ataaatatat	ttttagttag	ttatatatga	taggacaata	tcataaaaact	2340
cactttgtgc	catatcggtt	aacttgcgtg	agcttcaccg	ttaatcgata	aaaaatatta	2400
aaaaagcatc	tccggtttgc	tgcactgctc	gagttgtgca	aaaaaaaggc	acgacggacc	2460
cgcgcgctga	cgcggtgagc	cgcaagtccg	caacggcgcg	gccgcgcgct	aggaaaaaag	2520
cctcgcccgt	gaagcgaact	ccctctcctt	cgaccttcgt	tcttccactg	cggcctgcgc	2580
gacccgtgca	gctgcgcgct	ccacctggcc	gcgctggggc	ccacacgcgc	tggcatctgg	2640
agcattgccc	ccggacttcg	cgcggccgcg	cgcagccccg	ctccccaccg	aaaagcgaag	2700
cgattgccat	ccccacgcca	cgcggaagca	caaggtcccc	ccctgcacga	tcagcaggac	2760
ctcgccacgc	cgcgcgtgga	gctgcgcgtg	cgcgtgtgcg	cttggaccga	cgcgcaacgg	2820
cctgcctcga	ccgcccgtgc	acgccactgc	tcatgcagcc	gtccgcctcg	cccccgcccc	2880
gaactgccga	ggtcgcgtga	acgcccactc	ccctcacgcg	tcgtctccgt	gctatatagg	2940
cagcccgcgc	ccctccta	tgtagccctg	cagtcaccca	gagcagacc	ggatttcgct	3000
cttgccgtcg	ctgggggttt	agcattggct	gatcagttcg	atccgatccg	gctgcgaagg	3060
cgagatggcg	ttccgggttt	ctggggcggt	gctcgggtgg	gccgtaaggg	ctccccgact	3120
caccggcggc	ggggagggtg	gtctagtctt	ccggcacacc	ggcctcttct	taactcgtaa	3180
tgatcctgca	actcctcctc	cctctctgat	caagtgtggg	cctgattcgg	gtctgtatgc	3240
gagtgttgtg	gtggtgaact	ggtgaattgg	tgatgcattg	aggggtgtct	cgagttggat	3300
gttcggggac	gcacggggcc	atgcgcgcgg	cggccgcggc	caggaaagcg	gtcatggttc	3360
ctgagggcga	gaatgatggc	ctcgcatcaa	gggtgactc	ggctcaattc	cagtcggatg	3420
aactggaggt	tctgcatcca	ctcgtcactt	tcatgcattt	tatcacataa	ttcacctgaa	3480
agtctacatc	tacatgcatt	tctgattttt	acctcttttt	ggatgctatt	tgagaacaat	3540
gagacacacg	attagtgaga	tgcacaaacg	gttaaacggt	ccatggagct	ccacaagtct	3600
gtactgtacc	accatttaga	gttgctccata	caatcgtctt	ctgaacattt	cgtctcttcg	3660
gcacgggttc	aggtaccaga	catttctgaa	gagacaacgt	gcggtgctgg	tgtggctgat	3720
gctcaagcct	tgaacagagt	tcgagtggtc	ccccaccaa	gcgatggaca	aaaaatattc	3780
cagattgacc	ccatgttgca	aggctataag	taccatcttg	agtatcggtg	tgtattactt	3840

gcttctactg	caccactact	gaaacatttc	aagacgttcc	cgtctgtcat	gatatatata	3900
aggggtgtgt	tgggttggga	gtgagggaga	ataaagtgg	tccattctaa	ttttctggac	3960
gttttagttc	caacaaaata	agagtggagc	ggctcctgga	tctccatag	gaaatttact	4020
ataagtagtt	ggaacgctcc	cgtccacaa	aaacgacgga	tgcgagcgct	ctctcggaag	4080
tgagcgctct	cctggctctc	ctccctcta	cactcatag	gctctcaaac	caacaaaga	4140
atggaatgac	tccgctctat	tctcgtgaca	agattctcgt	gttatcttaa	ttcttgcata	4200
cttttaggtg	cgtttggttg	gggagccact	gggatggagc	ggctccattc	cagtttctaa	4260
aaattgagcc	atttcattct	gtgtttgaca	agcagaacgg	aaccacttca	tttttggttt	4320
ggttgaagag	tagaatagag	cggagccgct	ccatttattg	tttggttggg	gagccatag	4380
actgttgcgt	ggaggagagg	gatgagagcg	cttgcgctcg	cgagagcgct	cgcatcggat	4440
cgttttgggt	gagcgagagc	atcccaaata	tttgctattt	atagactcca	ggaaccgctc	4500
cgtctctctc	catggttaacc	aaacatcaaa	aaaataagaa	tggagccggt	ccattcccct	4560
ccgctcttca	accaaacaca	ttagtccagg	tcttcatagt	gttgctggaa	ggcatatga	4620
atggcttagt	caggttcggt	ttcacttgaa	atctttcata	ctgctttgat	gcctaaagta	4680
ttccttatgg	tttcaccaag	cattacaatt	gaacggacaa	gatatgttct	ctttcagagg	4740
atattttacag	aaatgtttagc	cagacaatat	agaggcaaca	gctgggttcta	attctttggc	4800
aaccttcatt	tccatccctt	ctgtattctt	cttataagga	ttcctatata	taaccatgc	4860
gaacaccact	atccatctca	ctgtgctttc	aacttccttg	aataaaagaa	caggtacagc	4920
ctctatagaa	gaatccgttc	agacattgat	gaacatgaag	gaggttggg	agccttctcc	4980
cgtagttatg	agaagtttgg	atttaatcgc	aggtattctt	taacatgaag	tgtttatctt	5040
ttatcttata	tactagcaaa	taactttgta	tcttattttc	caatgtagcg	cggaagggtat	5100
cacatatcga	gaatgggctc	ctggagcatt	tgtatgatct	ttcttcaatt	attctaattc	5160
tattgcttat	catgatacaa	tactagtctc	atgtttcatt	atgagaatga	tcactctccc	5220
agctacgaag	ctactcatat	ttaataaatc	ttttaccctc	aaaatatata	cattgtcaga	5280
cttagctcct	agtgttatte	agaattgact	cctgctttct	atcttagtct	gcagcattgg	5340
tgggtgactt	caacaactgg	gatccaaatg	cagatcgtat	gagcaaagta	tgcttatgcc	5400
ttcagtaaac	attatatatt	tcacttatta	gattggtttg	tttatctttt	tcattaactc	5460
aatatatgtg	cactctattt	aatcaacttt	ccaaccggaa	agtgcataat	ttgcattcac	5520
tggagtcagt	tcttacaatt	gactctgcat	ataagtgtgc	tattagatac	ttgtgttgct	5580
gctgatactc	ttagaagtca	atgtttccaa	actgcacatg	tgatacctgt	attctaattg	5640
ctcgatatct	ttacaaattg	catgttccta	taataggatg	aattgaatac	tcatacttag	5700
tgtgcatgga	cttgagtag	gtctaatctc	taatgattct	cgaagagttg	tttgtttgta	5760
agggttctta	cccctttatt	tcttcttaat	ataataatga	tacacaactc	tcttgtagtg	5820
tagcgaaaaa	aaaagaaaac	ttactgatcc	tcgcatagtc	taaatcgagg	tactttgttc	5880
ttgcagaatg	agtttggtgt	ttgggaaatt	ttctgacctc	acaatgcaga	tggtacatca	5940
cctattcctc	atggatctcg	tgtaaaagga	gccgctttac	ctcattgggt	gttgtttttt	6000
gctggggagc	atctccaaaa	ttaactgggtg	tgttttgacc	atataatagg	gagaatggat	6060
actccatcag	ggataaagga	ttcaattcca	gcctggatca	agtactcagt	gcaggcccca	6120
ggagaaatag	catatgatgg	gatttattat	gatcctcctg	aagaggtcct	cctttcccat	6180
atttctgtac	ttacaaattt	taaataatcc	tttgttccta	ctttaggcta	gcaatttggt	6240
tcttggaata	taccgtttat	ttcattctgg	aatttcttta	cacccctttt	ggatccttgg	6300
aattgaattc	catttctaata	atagtaattt	aggcaaatga	caattaagct	aatatggttt	6360
tatgtgaaat	atatttgtat	actattatta	gcaatatgtg	ggggatattt	atgtgctaca	6420
ttttactata	gaggagttag	ccgaagagcg	tcttataatt	tgacagatag	aaacatatta	6480
tggtgatata	taaaattatt	tctcatactc	cacctatga	atttgagata	ggcttatatc	6540
attgcttttg	aaagtggtaa	aatgtttaat	tccaagctaa	atagactact	ttattaagta	6600
aattccaatt	cctccaaaat	gaatagatct	aaaagagccc	cttagtgat	ttcagttagt	6660
cttattttct	gtggatctgg	attttacttt	tctcacttgc	tactaattaa	gatattttac	6720
atgcatatgc	aggtaaagta	tgtgttcagg	catgcgcaac	ctaaacgacc	aaaatcattg	6780
cggatatatg	aaacacatgt	cggaatgagt	agcccggtat	gtcacctttt	ccttttgatt	6840
cttttagacac	tgctaggcat	atgtaacacc	cggccactgg	ggacgtcgtc	accccagggc	6900
gagttttttt	ttgactgcta	gccacctgca	gttttagtta	attataggct	tcaacgatag	6960
ttagatacta	tcatcttctc	acattggctc	tccttatcta	agcgtatgta	atctaagtgt	7020
atatatttat	tttcataaat	gttcactttg	ctactttggt	agtgaagcca	gcaacattta	7080
attccaagtt	attctggaat	togtttaatg	atgttgtctt	gattgattga	ttccaatcgg	7140
atgtttgaac	ctctcaaaaa	tggatgttcc	aaaggagaa	tttttacatt	ttcttgcaaa	7200
attagtaaca	attttgaact	gattaatgta	actttgaaatg	tcatttctgc	tccgctatat	7260
gtttctatag	ctcatagcat	tatataacat	ttcaaaacaa	tattcatatg	ttgctcatga	7320
atggctcatca	atatgtttata	tttcttctac	tttagttgat	acattttgct	taatatattaa	7380

aattaaattc	atatttccca	caattacttt	gagttaaaaa	aaataaacat	ggccatagca	7440
cagccacagt	attgttactg	tcaccctaca	tttgtttcta	attgccatac	gattttacat	7500
acatgttctt	tcactgatca	ttatcctttt	ccggctatca	ttttgttata	ttgtttgcac	7560
tagataattg	cctgcattct	cttgtgatta	ggaaccgaag	ataaacacat	atgtaaactt	7620
tagggatgaa	gtcctcccaa	gaataaaaaa	acttgatac	aatgcagtgc	aaataatggc	7680
aatccaagag	cactcatatt	atggaagcct	tgggttaattt	caggatccag	ttttgtttgt	7740
ctttttcttc	tatatccact	ttatacctac	tgggtacctgt	tgagtgtttg	ggaaccttcac	7800
ctttcatcgc	cctctttttt	agtgtcattt	catgcctctt	tcactgtttg	ttctagtgtt	7860
atagcattaa	atgtacattt	tgggtattgtc	ttcgataatc	tattgtattc	agatgcatgc	7920
atagtaacaa	gactacacag	acatatatgt	aacgtcgtaa	cctcaacttt	atgaacatgc	7980
tctgattgga	ttgatgattg	catgatactg	tcaatatggc	attacagaaa	ctcgcaagta	8040
tcagtaccaa	tgcttatagt	ctcctgtcat	aaaacaagt	tgtttaggac	aagatttagt	8100
caaatcataa	aaactacaaa	tatcattatc	ttttaactta	tttagtttga	aagcatgtaa	8160
gttgacacta	aaaaaagcat	gtaacttata	tgtgtggaat	tgtattgaaa	aatccttgca	8220
aatctcatgc	agtcgaagga	ttaagtggtc	aaagatattc	atcaaagacc	atgcaatgtt	8280
gaatcaagta	ttttgatagt	aaggagtaca	tgtttccaac	agcatgaatt	ttgatacagt	8340
atccatctgt	tgagacgacg	gagatgatac	catccatctt	aaatttttagt	tccgaatgat	8400
accttttcaa	ttttgacaaa	caacatgaca	acgtcgattt	aataattttgt	tacttctatt	8460
gcagatacca	tgtaactaat	ttttttgcgc	caagtagtcg	ttttggtacc	ccagaagaat	8520
tgaagtcttt	gattgataga	gcacatgagc	ttggtttgct	agttctcatg	gatgtggttc	8580
ataggtaatt	gataaattct	ttgttaataa	ctcatcttcc	tttaatacaa	gtatcttatt	8640
gcatagttaa	agctaaat	gatgatgtgt	aattcatagc	tccaagaaaa	aggaaatata	8700
gtaagccttc	atgcaaaatt	gaaacattta	accctcgact	tactggaccc	gacctgcaac	8760
taatctacaa	cctctacaac	atacatcaga	agatgagaaa	taatcgccgt	aacataacaa	8820
ttaccctgac	taagatatat	gacaagtcaa	aatttttagct	gtaatgttca	gtgttctagg	8880
cttaagcttt	aacatcagga	aggaggggaa	gattatccac	caagtcccaa	atcaagaata	8940
ttcagaaaga	acctgtactg	tcaaagtccc	tttgatatat	gaaaaaactg	gacatgcgtg	9000
gggtaagaca	gtcccctggc	atttatattaa	gaagacctca	cacaggtcga	gaaaaacatcc	9060
gaaccgtgcc	acaccctac	acagcggcac	cgtagcccat	gtgagaaacg	accgcgtccg	9120
ggaccggacc	gttagatagt	gctttggtat	gtcatgagtg	aggggatatt	tttaaccaca	9180
gcctgaaatt	cgctcccacg	gtggtagaaa	accctcggt	gcgcagatta	taggagtggg	9240
tatcaagttg	gctggactag	cgggagcgaa	ggggcgccgt	gctgccgctc	ggctgtgaac	9300
tcagggtgagt	tgcggaacac	aggtgaggca	gggaacaatc	acttcttaga	ttgatcaatg	9360
cttcccaatt	gattacagac	gcctgtgact	ttatagtcca	actctaaca	gatccttata	9420
tcaggaactc	taacagattc	ttattaaaa	agactcctta	atctcaggat	ctctaaccag	9480
actctcctta	tctaaggatc	ctcctaactc	ttatctatat	gtccctctat	aatgggccga	9540
ctagagggct	gtttaggcca	cctactataa	accataaac	acggggagtc	aaacccagga	9600
tctgaggagt	gctactcaga	ccacctaac	aactcggcta	gaggcccttt	cgcaagcccc	9660
tttgatatca	gccacccaca	tcttattttg	aagtgttgga	catactgttc	gctgtttttt	9720
ctgctctttt	tggaaacca	ttgatcaagt	ttggaccaa	ttctgccatg	gtcttgccat	9780
taagccagcg	atcgaaacca	cattgagttc	tttaaaattt	agtttaacaa	acacttaatt	9840
ttagtttgat	tcataatttct	ttcttgaa	agcaggatag	ctgtgtatgt	ttgtattaac	9900
agagagaaat	gcaaacaccc	cactttgtct	tccttaggag	aacaaggggg	gttatggctg	9960
aggtttagag	aaacgtacaa	tcacctgtaa	ataggataat	atcaatacaa	tgctacttta	10020
aactagcccc	tcaacccaac	ctagtccatg	atcctggagt	ttacaggatc	cagcaaagtc	10080
ccgaagtttc	agatatttcta	agagcaogcc	atcttgcat	ttggactcaa	gggaagagtt	10140
ccctcgaagg	ctcacatttt	tttttgtctc	cataagggtcc	aagcccctag	aattacagcg	10200
ttgttgaaac	ccttcttttg	aagcatcaat	ccaatccgca	aagttctgct	cattttgtct	10260
tggaaaccaca	caaccaagac	caattggagg	ctcacatttt	ttttgtctcc	ataagggtcca	10320
agctcctaga	attaaagcgt	tgttgaaacc	cttcttttga	agcatcaatc	caatccgctg	10380
cgaagctctt	ctgcaagcta	accgcttttc	tgaagagaaa	tcaactattt	tctcctcgct	10440
tgtgaattcc	tatagatagc	agatagttct	tttcctgtg	ctttagttag	tttcttctg	10500
gtttatgtct	tgggagcgac	tgctaattgt	ctttgtccgt	cttctttata	ttataaaata	10560
tctaccacag	caggggcttc	tcctgtgtga	cattttcaaa	aaaaaacata	gaatgccaaa	10620
attgttaggg	aaaaaacaaa	atagtaagaa	tactgtttt	aaaggcgctg	cctaggcgctc	10680
caggcgcccc	aacactccta	gacgtccccg	tcgcctagct	aaatcagcaa	gagggcagag	10740
gggaagagga	aggggtggcg	ccgaatcggg	gccggaggcg	gcgaatcggg	ggcgacggcg	10800
gtgagatcag	gggaggcggc	gaccaaactg	tggccaacgg	cagccaaacc	tagggatggc	10860
aacgttttta	aaacctgcgg	gtagaggggt	cacaaaccca	caccgcggg	tctaattatta	10920

aacccgcacc	agtaccatt	acccgtgacg	ggtataatat	gatacccata	ccctctaccc	10980
gcgggcatat	aattatacac	atatatTTTT	atatatcatc	tgtgtatgca	tatatTTata	11040
cacacatata	aatatgcata	cgcacgcacg	acgcacgcac	gcgggtttgc	gggcacgggt	11100
acaacatttt	cataccgcg	agaaaaaac	cgttgggttg	agaactaaac	ccacacccta	11160
tggggTTTT	acccgcggc	acgtgggtta	aatgtgtccg	tgccatccc	tagccaaacc	11220
gagcgcggcg	gcggcaaaat	cgtgggtgac	ggcggctggc	aggccacatg	ggctgggctc	11280
ttatgcccat	acacacaaac	aacacacaac	agtcgaacac	tcctcttctt	gttaatctaa	11340
ccatagtaca	atatttgttt	ttattagaaa	gccaataaac	actttgcatt	taaatagcgc	11400
ccaagatttc	tacactcact	tccaggtcca	aaggtgagcc	atggagaaag	gctctatgtt	11460
gatttgaac	ataaggagcc	actttctcat	gtttccatat	atgtggatcc	tgatccatgt	11520
tcacagatt	gattcatttt	ctaaggacat	aaagtcatag	ggtttcccta	ttccttttgc	11580
tacagggccc	ttgtgattct	ggtttctgaa	tgtgtgaatt	gtgaaagggc	ctctagtga	11640
gttggttact	caatccggtt	gctgcactag	caaatcagcc	tactctactt	tctttgagg	11700
gtctgtcaag	ttcagacctt	gcagattaat	ttggaggagt	tgggctcctt	tctgctgcaa	11760
gttctttctt	tggcttgttg	ttgctgcaat	agcaaaccca	tggttgactg	cggatccgct	11820
tgccaaacga	ggccttcctc	acccagattc	ctgctccttg	tgtgactaag	ccggtgagac	11880
cattcagcac	attctaattg	gttgtgtctt	ctctagacaa	atttgggtct	tgactttgca	11940
atccttgcac	ctggaggcca	ttacgcccaa	ttggacagag	ttgggtttct	tctcatggtg	12000
ggcgcgctct	ggcaagatgg	ttcccaagaa	cattcgaaag	gggttgaaac	ctctttgtat	12060
tccgatagcc	tgagaaattt	ggaatttagg	aacttgtgtg	tgttcaaaga	ctcccaacca	12120
aatgtacaac	tccttctaca	taggatagga	agtgagggtc	ttctttggtg	tgctgcaggg	12180
ttgcttgagc	tcctgaatag	gtcgttgtcc	ccagcctgct	aggtgctggt	ggccttggtt	12240
gtcagctttt	gttgtgtttt	tgtgtaaaac	ctagctggct	agctgttttg	actagggttg	12300
ggtcattttg	atccttattc	ctgtcttttt	cttctgtaaa	taaaatgaca	cgtagctctc	12360
ctgcgtcgtt	cgagaaaaaa	agtactcgtc	agaatagaaa	attttactct	ttgataaact	12420
ataactacct	aatcaagaca	agaatataaa	attttaaata	ttggaaacct	atttgtcgtg	12480
gttcaagggt	ctttggtctc	caagcttttg	gatgagatcg	attataatct	gttatgtaca	12540
accttatgat	tatttgaagc	tctttgggtt	cataccttta	actttgcttt	tgtgttactt	12600
gcagtcatgc	gtcaagtaat	actctggatg	ggttgaatgg	ttttgatgg	acagatacac	12660
attactttca	cagtgggtcca	cgtggccatc	actgatgtg	ggattctcgc	ctattttaact	12720
atgggaactg	ggaagtacgg	aacaaaaatg	ctctatctcc	attaaatttt	ttctctctatt	12780
tttctgcctg	tatcgttcca	acaattttat	ccgtatgcag	gttttaagat	ttcttctctc	12840
caatgctaga	tggtggctcg	aggaatataa	gtttgatgg	ttccgttttg	atgggtgtgac	12900
ctccatgatg	tacactcatc	acggattaca	agtaatttaa	gctttatgcc	tgttagttaa	12960
tcttcacttg	ctaagtctga	ctggaatact	ggattatgcc	tgggaactag	ttttgtttag	13020
tatcatattt	gttatatatc	attccttctt	ctaatctaaa	gtcatgcatt	ttacttttag	13080
taacattttac	ggggaacttc	aatgagtatt	ttggctttgc	caccgatgta	gatgcagtgg	13140
tttacttgat	gctggtaaat	gatctaattc	atggacttta	tcctgaggct	gtaaccattg	13200
gtgaagatgt	aagtgctgag	tttgctgtgc	atttaatatg	aattctcgca	tatatTTgtg	13260
gaaatatttt	tgtagtcgaa	gttgcttttg	tttatctaga	caagatactc	ctatttggtt	13320
atgcagaagt	taatttgaat	tttaatacga	agtgcacact	aagttactgg	ttaatatTgt	13380
tcttcatttc	ttcaagtTtc	agtctatttc	aactccattt	ataataatgt	gcttggaag	13440
ttactgtttt	aattttacat	taatacacaa	tacaacaaga	tcatgcttta	caacatgtgt	13500
gtatattaga	taagtgtTtc	atgcacatag	agttgctact	ttttgaagaa	acatatagaa	13560
ttacttaaaa	ggaactatTt	gaaatacatg	aagaaagtta	atggctgaac	ctatatTgta	13620
atggacagaa	cactagaact	ttcgggtTta	tactgagcct	ggactataaa	gaagattTct	13680
cttgaagtTc	aatgagtTct	cgaactaaat	attcctttct	aactactgga	tggataccaa	13740
agacacaaaa	aattgaaatt	gtaagccact	ctgctctTgt	tttggcacca	gatttaatat	13800
aatataaaaa	aaaatttaata	atggggagaa	gactccctt	ggattggctt	ttaaagagag	13860
agtacagaac	ccttcccaac	tcgtccccga	accaaacttt	gccaataatc	ggatccgcct	13920
cgtcctgtcc	ccagtgtccc	ttggaaccaa	acacatgctg	agggaatata	ctcctggggg	13980
atggagttat	ggacoggttc	caactcgtct	ccgaaccaat	gatttcaagt	cgtccgacta	14040
atcgcgatta	gtcgggctgg	tcggcaatta	ggacacgatt	tgctgggcca	ctcgactaga	14100
agacctagtc	gtcctggctg	ttcgactaat	cgctcgactag	ggcgactagt	cgttcgagtt	14160
atgtgtcctg	gtcgtccccg	ctagggtTta	gtttattggg	ccttttttag	cccatctaca	14220
gtctaattta	gagtaccctc	tcctctcctg	ttctaacccta	gccgccaaca	gtctccaca	14280
cctacacgtc	tcctcttctc	tcctcccttg	gcatgccctg	cagtcctgaa	tcctgattgc	14340
cggcgactac	accggcgggt	acagcaccct	gcgtcccttc	agggccggcc	ctggacaggt	14400
gcggcggtg	cggcgcgacc	gggcctccga	aaaccagagg	gccccctccc	gtgtatacgt	14460

gtatgtgtat	atatatagaa	aatgtttatg	taaattacaa	cacaaagatt	gataaaaaggc	14520
caaagcccat	caacgtacaa	tatgcatgcc	tgtttaaact	cgattacgta	cttggcttct	14580
cgtggacgct	tgagctgcaa	tcctctttgc	cttctcttgg	aagcctccgc	tcatgttgtc	14640
caggctacaa	caattgttcg	gcaccagcag	tagcagctag	cagcgacaga	cgggagggat	14700
catcttgctc	gttgctctgt	catctgcccc	aaccaggctt	gccgcgcgcc	gtgcctagcg	14760
gtagcgacat	ggtagcgtgc	cctgcctcgg	taacttcgtc	aagtctgcta	ctctactttc	14820
gttgttcaat	tttcgcctca	aaccaatctt	tactagttaa	atgttctttt	atctctgagt	14880
ataggttatt	ttgtgttctg	atcaccacat	tccacatgca	tctagttagt	aaaaaggaaa	14940
aaaacataaa	gataagttca	atgcatcgta	gtattcatat	ttttttggta	tttgtgatga	15000
atggggcctt	aatttttgtc	tcgcaccggg	ccaccaaatt	ctcaggggcg	gccctgcgtc	15060
ccctgccccct	tctcccgag	tctgtctcgc	ccgtgcact	gctgcagtct	ccaaccgatg	15120
gtggtgcga	gccagtgcag	agatgagcat	gtcatctgaa	agtccttcct	gtccccact	15180
tctcccccta	aaccctaaat	aaccctaaac	cctaagcaac	atgttacttt	atctatccta	15240
gttggtttgtt	ctaatttata	taatgtatac	aggtatattt	atatatatac	ctatataagt	15300
ataactacta	gtctaggacg	accagggaac	gactaggggt	cgactagtcg	ccctagtcgt	15360
cgcctaactcg	cgactagtcg	cttggtcggg	cacagagttc	gactaggcga	ctaggcgact	15420
tgaaatcatt	gctccgaacc	aaacactcag	gatcaccttg	tcctgtcctg	gtggtccttg	15480
gaaccaaaca	cacatttagg	gaagtaatca	tctgacctgc	aagcacctgg	gttcatgcag	15540
gtggatagga	tctcagctag	aggtcttacc	aaccaagtta	ttgttcgatt	ctcagttctc	15600
atgctccggt	gtgcaatcta	atgtaaataa	atgtaccacc	gtgttgtagc	catataccag	15660
cgtaatctcc	aacaagggct	gtcatgtgga	atcaatagag	gtctctgtca	taatctgcaa	15720
cggcaacaaa	ggacatacca	atctttaaca	cgaagagaat	taactccttc	ccatctgtag	15780
ttagatatta	gaatattaga	attgcatatt	tttatgggtt	cttgaaggaa	gtggcaaacc	15840
tcaatattga	actctacca	cacatgtata	atatgaaaaa	taaagttatg	gtatatttac	15900
gagtaaatat	ttctatgcct	tataacggga	ggaaaaatat	cgaattttga	tattgactta	15960
ataggacatt	gtcacgttat	tgtcaatatt	aacttatatt	atggatgtca	tcacaaaata	16020
actatacctt	ttagttttta	aagaaagggt	aaaacataca	tatatattga	acgtttccgt	16080
ccattttcca	ttaatgtttg	attgtttttt	tctctcttgt	tgcaacatag	tgacatgttt	16140
gctagtttga	caaaatttag	gcaagatcat	tggcttacat	aaaccaatag	gaaatttgaa	16200
tttatctaaa	ataattttgt	agttttcttg	ttttctcaa	aattatataa	cttgttcagg	16260
ttagtggaa	gcctacattt	gcccttctct	ttcacgatgg	tggggtaggt	tttgactatc	16320
ggatgcata	ggctgtggct	gacaaatgga	ttgaccttct	caagtaagtg	tttcatatgt	16380
atgtggacga	tcattatttt	gttttattgg	gctgcttatt	taaataatta	ttttttttgg	16440
cataaatgtt	ataagtacat	acactaatgt	ttaattgcat	gggccatgtg	catatgtatt	16500
ttttgttga	tttgtcacca	cttttgggt	ttactttaca	tagtcataat	gagataatgt	16560
tttatgcaca	tttgcattgt	gctgcactta	ttgataacga	cacaaaatgt	actcgagatt	16620
gatgtatatt	ttcaaacttg	aaactaatga	acacatatga	tatacattgt	acactccatg	16680
ttacaaatta	taagacgttt	ttgcattttt	agatatattt	ttttactatg	tatctagaca	16740
aagtgtatat	ttaagtgcac	tgcaaaaggca	atgcatctag	aaaagccaaa	atgcctttgc	16800
atttggaaacg	gaaggagtaa	tattgttgca	caaactctga	agttttctat	gcataggata	16860
ttgcacatac	cagcaatata	tcagtgtgtc	atatatatatt	tgtataacta	acattacacc	16920
attcactgtg	ttcatatggt	ttgtgtttta	ttttctatac	aaaaaatctc	ttgaatctgg	16980
tactcaaate	atttgtataa	gcttttgggt	caaacgggca	acgctctatc	tgtaatatgc	17040
catctgcgta	aaagacaagt	atgtaatttt	gtgatccttt	ggttgtgtgg	ttgtcgggtg	17100
aaccgggatt	ctttaacctt	ttatccttct	tttatataat	gatacacaaa	ctctcctgtg	17160
cgttcgagag	aaaaaatgcc	ttctgcattc	actttgagat	atgtggtgag	tttcaatttc	17220
tatttaaccg	cacaggcaaa	gtgatgaaac	ttggaagatg	ggtgatattg	tgcacacact	17280
gacaaatagg	aggtggttag	agaagtgtgt	aacttatgct	gaaagtcatg	atcaagcatt	17340
agtcggcgac	aagactattg	cgttttgggt	gatggacaag	gttaccctac	tcattaatct	17400
tcttggtgta	cttattggga	catagatcat	gtttcacgta	ttgtttttta	caatgattaa	17460
ttctattttg	ttcttcaagc	tcaaggtgta	ttcatgggtc	ccacaaaaca	aatgttttat	17520
taaaagcaaa	gagttagaac	ttttgttagt	tttctttaat	ttggacttgt	gtcactgttt	17580
ctctgtcatg	actcatgagc	attatagtgt	caatttcacc	taagtgaagt	cctgtttttg	17640
gacagttcag	agtgaactac	gacttattgt	tttaataact	cattgagttt	gtagaacaag	17700
tattggctct	ttctcatcct	atactttcaa	aagtaggatt	tggccttttt	cttcactat	17760
tctttgaaaa	atcacatgtag	aaacacggaa	caataataaa	tggtaacatg	agaacctcac	17820
tggttctatt	tatgcaggat	atgtatgatt	tcatggccct	cgatagacct	tcaactccta	17880
ccattgatcg	tgggatagca	ttacataaga	tgattagact	tatcacaatg	ggttaggag	17940
gagagggcta	tcttaatttc	atgggaaatg	agtttggaca	tcctggtgag	atttaactac	18000

tttgtttcat	ttaaccttcg	ttgagtcctta	tagaacagta	cctcatccaa	caattatctt	18060
gcaatttatc	ttttgttagt	tatattagtg	ttgaggactt	gaggtcattt	tcttcttatt	18120
attttgcaga	atggatagat	tttccaagag	gtccgcaaag	acttccaagt	ggtaagttta	18180
ttccagggaa	taacaacagt	tatgacaaat	gtcgtcgaag	atttgacctg	gtaaactttc	18240
tttgattgtg	caaaagtcca	agtttgtatt	tacttttacc	actgatccag	tgctttaatc	18300
agcaaggtgc	cattataata	gttccttctt	tattcatatt	agcatgttcc	agaagtaaaa	18360
atattactac	ctttgtaaaa	gttttcttta	atataatgtc	ctttgtttgg	tcaataattc	18420
gtcattatcg	gaattgtgtt	atttttacat	tgtcagggtg	atgcagacta	tcttaggtat	18480
catggtatgc	aagagtttga	tcaggcaatg	caacatcttg	agcaaaaaata	tgaagtatgt	18540
tcttttttta	cttttttgat	ttggttctgc	aagggttcca	caaacatcat	atttgttgtg	18600
cattctactt	gtaatgtcat	tttaaaaaaa	atcattcttc	agttttactg	agcttttaag	18660
caatgaaggt	ttcattatga	attctttcat	gttgcatcaa	caactcttag	gtattttcat	18720
gatcattaat	agtactctgg	agacagcacg	ccataatggg	aacgaaaaat	tttctgatga	18780
aatttgctgt	gtaattgcag	ttcatgacat	ctgatcacca	gtatatttcc	cggaaacatg	18840
aggaggataa	ggtgattgtg	ttcgaaaaag	gagatttggt	atttgtgttc	aacttccact	18900
gcaacaacag	ctattttgac	taccgtattg	gttgctgaaa	gcctgggggtg	tataaggtat	18960
gcattctatc	tgcatctcct	atgctcaaag	tgcatctctt	ttcttgataa	atgagttaga	19020
tatacgtact	atcatgctgc	aatttatcaa	gtgtcattat	tgatctcttt	ctacggtgaa	19080
gctaggagca	gctaagctgt	tggtgtcagc	aattcatgtt	gtagttaatt	taatttgctt	19140
gaaaacgtag	gacgctagat	ttggattttt	ccaattttta	ggctgcacga	gcaggtaaaa	19200
ggtagcaaaa	tactagggcg	ccatgtttac	atgtataaaa	aaaacaaaac	aaaaaagaac	19260
taggagttcc	tgtgacggat	agccgcatgc	tcgttctctt	ggcgtctctg	atattggagc	19320
acatccgttg	tctgaaaaact	agccggagaa	gtttctcaaa	atcccactag	cggaggtgct	19380
gacagcattt	gctatttgta	ccaggtggtc	ttggactccg	acgctggact	atttgggtga	19440
tttagcagga	tccatcacgc	agccgagcac	ttcacgcgcg	taagttttgt	ggcacgtgat	19500
actgctctag	gtacgcagat	gtccacttgt	tcctgacaga	ggtgaactaa	cttctgttat	19560
ggccattact	tgcaggactg	ttcgcatgat	aataggccat	attcattctc	ggtttatata	19620
ccaagcagaa	catgtgtcgt	ctatgctcca	gtggagtgat	agcgggggtac	tcgttgctgc	19680
gcggcatgtg	tggggctgtc	gatgtgagga	aaaaccttct	tccaaaaccg	gcagatgcac	19740
gcattgcata	tacaataaag	ttctgatact	ttaactcgat	ctggaaaagcc	catgcatctc	19800
gctgcgttgt	cctctctata	tatttaagac	cttcaagggtg	tcaattaaac	atagagtttt	19860
cgtttttcgc	tttcttaatg	cttgatggct	gattgtttgc	acttgtttca	ttccgttggg	19920
cactgatggg	cttagagtta	gacaatcggc	tgcagcgcac	aggtttcaag	ctgggggggtt	19980
gcattgtccg	ctacagagca	gccagcaatg	tgccccctgc	tgccctgctc	gcttactttt	20040
aaatgccacc	cctcccgatt	accgactcac	tcagattcag	acacgcaaag	cactaccttt	20100
ccagtgtccc	tgaagacact	actccccgtc	ccagacttgc	cgttgacagt	atacctcgcc	20160
ttgcctgcct	tagatgttaa	tcctacaacg	atagacatgg	atacgggggtt	ttactaatgc	20220
ttggatgcat	gcattatcgt	atcctcgcct	cggagacatc	acgcgtgcat	ttggtcacac	20280
caactggtga	cagggaaagt	acactgagat	tgctcaagag	ttactcctac	tcgacttggt	20340
ggatctcacc	taacgctttt	caagttttta	tgatacttcg	ctggttggtc	tggccttggg	20400
ccagtccgga	gccgctctcc	tgtcacatc	acatgtacgt	ggaatgatgt	tgtctcgggt	20460
catggcatac	agcttggttg	gttttatctt	ctccatcgat	gccagggaag	cttgtgctat	20520
cattaccata	agcagctaatt	agtatgagtt	gatgagtgc	aagtgcacac	tgactaaga	20580
tacatataac	atcaatcaaa	acaatataac	aattcatata	gatcaccggt	gcttgcttaa	20640
aggttcaggc	ttcgagtagc	taggacacta	gctagtactg	ggctctcgat	cgtttgcaaa	20700
acaatacaac	agtaaaaggg	agagcgagca	catggaagct	ggaggataaa	atagatggac	20760
ggttggtaga	tgcaggetac	tcctactacc	aatttcttta	cacggcttag	ggcgcaagag	20820
gcctctcctg	ctctacacac	tgccctagct	gctagaccg	tttgaggggc	agtgttgcca	20880
gctagcagcg	acgcagagatg	gagtgcagcg	gcaaggagga	gcagcagatg	cagatcgtgt	20940
gcgtgcgcag	cgctgcgacg	ggcggcggcg	aggaggtggg	ggagtgggcg	gagcagtcgt	21000
cgcggtcggc	gctgtcgctg	ttcaaggaga	aggaggagga	gatcgagcgc	aagaaggtgg	21060
aggtgcggga	caaggtgttc	tccatgctgg	gccgggtgga	ggaggagacc	aaacgactgg	21120
ccttcatccg	ccaggagctg	gagctcatgg	ccgacccac	gcgccgggag	gtcgacgcca	21180
tccgcaagcg	catcgacaag	gtcaacaggc	agctcaagcc	gctcggcaag	acgtgctca	21240
aaaaggtacg	tcaactacaag	acactggaac	acaacacgag	tgtttctcca	ttcgggtttt	21300
tcattcgacc	gaatggctat	agaacgaagt	caagcaggtg	gcaacacgta	ctgcttgctg	21360
atcgtgtgtt	tgtctccgaa	catgtgcgtg	tgcaggagaa	ggagtacaag	atgtgcctcg	21420
aggcctacaa	cgagaagaac	aacgagaaag	ccacgctggg	caataggctg	ttggaggtca	21480
gttcaattct	tttgcaatcc	gcgcgcgcta	tagctaagctt	ttgttttgag	aaccgtctcg	21540

10/13

```

gtaggcatat tttgttctgc tgcggaaaat aaaaaggatt aaggctgttt tagaaccaat 21600
aataattcct gtgaaattca tataccaatc cgatggatgg aattgggttg gccatttcca 21660
tattattact gtactactag gcccaactga cttgggtcctc tatTTTTTcc ctaataagaa 21720
cgtgtctgtt gacacatgtt tcactcagag gttgggtcatc taaaacaata aaataataac 21780
gtgttctttc tggcagctcg taagcgagag cgagaggctg cggatgaaga agctggagga 21840
gctcaacaag acgatagaat cctctacta gatctgatcc ttgggtcttg ggcttcaaag 21900
aagcaaaatt ccgttgctct ttctaccgcc gaagccgagt acattgctgc aggacattgt 21960
tgcgcgcaac tactttggat gaggcaaac cttagggact acggttacaa attaaccaaa 22020
gttcctcttc tatgtgataa tgagagtgc atcaaaatgg cggataatcc cgtcgaacat 22080
agccgcacta aacacatagc cattcggtat cacttttttaa gggatcacca acaaaaggga 22140
gatatcgaga ttcatatcat taatactaaa gatcaattag ccgatatctt taccaagcct 22200
cttgatgaac aaacttttta caaacttagg catgagctaa atattcttga ttctaggaac 22260
ttcttttgtt gatttgcaca catagctcat ttatatactt ttgatcatgt ctctttcata 22320
tatgctatga ctaatgtgtt ttcaagtga tttcaaacca agtcataggt gtattgaaag 22380
ggaattggag tcttcggcga agacaaagg tccactccgt aactcatcct tcaccatcat 22440
ccgagcaatc tccatctttg gggagaaaag gactccatct ttggtataat cttcactcat 22500
ttattttatg ccaaaggggg agagagtaat tcaagggctc taatgactcc ttcttggcga 22560
ttcatgccaa agggggagga agtattagcc caaagcaaaa ggaccgcacc accacctaata 22620
tttaaaagtt ttttttcaat tggatgatt ctcaattgat aaatttcaaa ttggtatctt 22680
attgtgttca aaagggggaa aaaagtagtt tttcaaaatt gatatgacaa caccctcttg 22740
aacactaaga ggagaatttc atctagggga gctttgttta gtcaaaggaa aagcatttga 22800
aacaggggga gaaaatttca aatcttgaaa atgcttcgca aaatcttatt catttacctt 22860
tgactatttt gcaaaagaac ttgaaaagg atttcaaaa agaatttgca aaaacaaaac 22920
atgtggtgca agcgtggtcc aaaatgttaa aatgaagaa acaatccatg catatcttat 22980
aagtatttat attggctcaa ttccaagcaa cctttgcact tacattgtga aaactagttc 23040
aattatgcac tccatattt gctttggttt gtgttggcat caatcaccaa aaagggggag 23100
attgaaaggg aattaggctt acacctagtt cctaattgat tttggtggtt gaattgtcca 23160
acacaaataa ttggactaac tagtttgctc tagtgtataa gttatacagg tgccaaagggt 23220
tcacacttag ccaataaaaa gaccaagaac taggttcaac aaaaagagca aagggataac 23280
cgaagtgtgc ccagtgttcg gtgcaccagg ggacttcacg ctgaactcct caccttcggg 23340
aaaatccaga ggcgcttcgc tataattcac cggactgtcc ggtgtacacc ggacagtgtc 23400
cgggtgctcca aggaagggcg cctcaggaac tcgccagctt cggaattc 23449

```

<210> 3
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> /note = "synthetic construct"

<400> 3
 gactgaattc ctgcgcagga ggcagagctt

30

<210> 4
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> /note = "synthetic construct"

<400> 4
 gatcgaattc catagatacg tggagcagca

30

<210> 5
 <211> 20
 <212> DNA
 <213> Artificial Sequence

11/13

<220>
<223> /note = "synthetic construct"

<400> 5
ggcgacacga ggcacagcat 20

<210> 6
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 6
ggtcgactcg agtcgacatc gatttttttt tttttttt 39

<210> 7
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 7
gactgagctc ataccaaatg aagccaggag 30

<210> 8
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 8
ggtcgactcg agtcgacatc ga 22

<210> 9
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 9
ccagctccac ggttggttcgt gt 22

<210> 10
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 10

12/13

cgatggatcc tgtgacggcg tgtgagtcctc 30

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 11
gatcggatcg aactgatcag 20

<210> 12
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 12
cctaattgta gccctgcagt ca 22

<210> 13
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 13
gactggatcc tcgccttcgc agccggatcg 30

<210> 14
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 14
gaaagtcgac gaagagagaa tgaaagcgaa 30

<210> 15
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 15
gcgcgggtcc gtcgtgcctt tt 22

<210> 16
<211> 30

13/13

<212> DNA

<213> Artificial Sequence

<220>

<223> /note = "synthetic construct"

<400> 16

gatagtcgac cgacgcgcaa cggcctgcct

30

<210> 17

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> /note = "synthetic construct"

<400> 17

gactggatcc tcgccttcgc agccggatcg

30

<210> 18

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> /note = "synthetic construct"

<400> 18

gacgctcgac cgctcgtctc cgtcctatat

30